Volume 35 Number 1 March 1984

BJPIA 35 (No. 1) 1–100 (1984) ISSN 0007–0882

The British Journal for the Philosophy of Science

Published for the
British Society for the Philosophy of Science
by Aberdeen University Press

THE BRITISH JOURNAL FOR THE PHILOSOPHY OF SCIENCE

Editor: DR. DONALD A. GILLIES

Subscriptions
The *Journal* is published quarterly. The subscription price for mailing addresses in Britain is £14.00 net per volume post free and for mailing addresses outside Britain £17.00 (or foreign currency equivalent calculated at the day's rate); single issues £5.00 net (overseas £6.00) plus postage. Orders may be sent to any bookseller or direct to Aberdeen University Press, Farmers Hall, Aberdeen, Scotland AB9 2XT.

For members of the British Society for the Philosophy of Science—who should order direct from Aberdeen University Press—the subscription price is £9.00 (overseas £12.00). Membership of the Society is international and is by no means confined to people resident in the United Kingdom.

Microform
The current volume of the *Journal* is available in microform to subscribers to the printed version, *at a combined print/microfiche rate of* £19.50 (*overseas* £22.50).

Back Volumes
Volumes 1–16 may be obtained from Kraus Reprint Ltd., FL-9491 Nendeln, Liechtenstein or Kraus Reprint Corporation, 16 East 46 Street, New York, N.Y. 10017. Volumes 17–34 may be obtained from Aberdeen University Press, Farmers Hall, Aberdeen, Scotland AB9 2XT. Prices on application. Volumes 1–33 are also available in microform from Kraus Reprint Corporation.

Typescripts and Review Books
Typescripts, correspondence and books for review should be sent to the Editor at the Department of History and Philosophy of Science, Chelsea College, University of London, Manresa Road, London SW3 6LX.

Review books are accepted on the understanding that, although they will appear in the 'Recent Publications' list, a full review cannot be guaranteed, and that books not reviewed are not returnable.

Notes to Contributors
Three copies of all contributions must be sent. These need not, on initial submission, be in any particular style, provided they are typewritten and clearly legible. However, contributions accepted for publication may have to be wholly or partly retyped to accord with the *Journal*'s usual format—details of which may be obtained from the Editor.

Articles should be accompanied by abstracts of about 100 words.

Material is accepted for publication in the *Journal* on the condition that copyright is assigned to the British Society for the Philosophy of Science.

© British Society for the Philosophy of Science 1984

For permission to reproduce material from *The British Journal for the Philosophy of Science*, please apply to the Editor.

CUK-H0685-23-P16285

The British Journal for the Philosophy of Science

Volume 35

Published for the
British Society for the Philosophy of Science
by Aberdeen University Press
1984

P 16285

CONTENTS

ARTICLES

page
1 THE DYNAMICS OF BELIEF AS A BASIS FOR LOGIC
Peter Gärdenfors
11 MELLOR, CHANCE AND THE SINGLE CASE
Huw Price
25 CONTEXTUAL HIDDEN VARIABLES THEORIES AND BELL'S INEQUALITIES
Abner Shimony
47 PROBLEMS WITH ELECTORAL EVALUATIONS OF EXPERT OPINIONS
Roy A. Sorensen
101 RELATIONISM AND POSSIBLE WORLDS
J. Butterfield
115 THE STANDING OF PSYCHOANALYSIS
Edward Erwin
129 MENO—A COGNITIVE PSYCHOLOGICAL VIEW
Benny Shanon
201 TRUTH, RATIONALITY AND THE SOCIOLOGY OF SCIENCE
Richard C. Jennings
213 THE CALIBRATION QUESTION
Frank Lad
223 IS CONSCIOUSNESS IMPORTANT?
Kathleen V. Wilkes
313 THE DUHEM THESIS
Roger Ariew
327 THE PASSIVITY ASSUMPTION OF THE SENSATION-PERCEPTION DISTINCTION
Aaron Ben-Zeev
345 INDIVIDUALISM AND GLOBAL SUPERVENIENCE
Gregory Currie
359 A REASONABLE REPLY TO HUME'S SCEPTICISM
Richard H. Schlagel

DISCUSSIONS

55 PROBABILISTIC CAUSALITY AND PREEMPTION
Douglas Ehring
57 ARE PARADIGMS INCOMMENSURABLE?
Allan Franklin
60 EINSTEIN AND SPECIAL RELATIVITY: WHO WROTE THE ADDED FOOTNOTES?
Ian McCausland
62 A GEDANKENEXPERIMENT TO MEASURE THE ONE WAY VELOCITY OF LIGHT
Charles Nissim-Sabat
149 BEHAVIOUR THERAPY: A RESPONSE TO EDWARD ERWIN
M. E. Grenander

page

156 THE PHILOSOPHICAL RELEVANCE OF 'TECHNICALLY GOOD' EXPERIMENTS
Tyrone Lai

245 BAYESIANISM AND SUPPORT BY NOVEL FACTS
Colin Howson

375 O'HEAR ON AN ARGUMENT OF POPPER'S
Peter Glassen

377 REPLY TO GLASSEN
Anthony O'Hear

REVIEW ARTICLES

65 AN UNREAL IMAGE (review of Bas C. Van Fraassen: *The Scientific Image*)
John Worrall

161 A BAYESIAN APPROACH IN THE PHILOSOPHY OF INFERENCE (review of Paul Horwich: *Probability and Evidence*)
I. J. Good

166 T. S. KUHN, FUNCTIONALISM, AND SOCIOLOGY OF KNOWLEDGE (review of B. Barnes: *T. S. Kuhn and Social Science*)
Homa Katouzian

253 POPPER, PROPENSITIES, AND QUANTUM THEORY (review of K. R. Popper: *Realism and the Aim of Science*, *The Open Universe*, and *Quantum Theory and the Schism in Physics*
Henry Krips

274 UNIFICATION IN SCIENCE (review of C. F. Von Weizsäcker: *The Unity of Nature*)
Michael Redhead

280 SPACE-TIME PHYSICS AND THE PHILOSOPHY OF SCIENCE (review of Michael Friedman: *Foundations of Space-Time Theories, Relativistic Physics and the Philosophy of Science*)
Roberto Torretti

381 THE EPISTEMOLOGY OF EXPERIMENT (review of Ian Hacking: *Representing and Intervening*)
Allan Franklin

390 NANCY CARTWRIGHT'S NEW PHILOSOPHY OF PHYSICS (review of Nancy Cartwright: *How the Laws of Physics Lie*)
Peter Gibbins

REVIEWS

81 JENSEN, V. J. and HARRÉ, R. (*eds.*): *The Philosophy of Evolution.* Review by Vernon Pratt

83 NOZICK, R.: *Philosophical Explanations.* Review by Leslie Stevenson

86 O'HEAR, ANTHONY: *Karl Popper.* Review by James M. Brown

90 VAN DALEN, D. (*ed.*): *Brower's Cambridge Lectures on Intuitionism.* Review by Joan Weiner

95 WEATHERFORD, R.: *Philosophical Foundations of Probability Theory.* Review by Peter Milne

175 KYBURG, HENRY E., Jr: *Epistemology and Inference.* Review by D. H. Mellor

179 MAGEE, BRYAN: *The Philosophy of Schopenhauer.* Review by Gordon Blair

page

184 Radnitzky, Gerard and Andersson, Gunnar (*eds.*): *The Structure and Development of Science*. Review by Robert Nola
188 Roseveare, N. T.: *Mercury's Perihelion from Le Verrier to Einstein*. Review by Derek Raine
191 Webster, Charles: *From Paracelsus to Newton: magic and the making of modern science*. Review by Simon Schaffer
293 Gärdenfors, P., Hansson, B. and N.-E. Sahlin (*eds.*): *Evidentiary Value: Philosophical, Judicial and Psychological Aspects of a Theory*. Review by D. V. Lindley
297 Harman, P. M.: *Energy, Force and Matter* and *Metaphysics and Natural Philosophy*. Review by Steven French
301 Katz, J. J.: *Language and other abstract objects*. Review by Mike Dillinger
303 Whiteside, D. T. (*ed.*): *The Mathematical Papers of Isaac Newton, Volume VIII: 1697—1722*. Review by A. J. Apt
307 Stove, D. C.: *Popper and after. Four modern irrationalists*. Review by J. M. B. Moss
403 Blackwell, K. *et al.* (*eds.*): *Cambridge Essays 1888—1899* by Bertrand Russell. Review by C. W. Kilmister
404 Hussey, Edward (*ed.* and *tr.*): *Aristotle's Physics Books III and IV*. Review by Sarah Waterlow
408 Lightfoot, David: *The Language Lottery: Toward a Biology of Grammars*. Review by Fred D'Agostino
408 Julia, Pere: *Explanatory Models in Linguistics: A Behavioral Perspective*. Review by Fred D'Agostino
411 Maze, J. R.: *The Meaning of Behaviour*. Review by W. D. Hart

MISCELLANEOUS

194 RECENT PUBLICATIONS (June)
415 RECENT PUBLICATIONS (December)
100 ANNOUNCEMENT: *PHILOSOPHICAL INQUIRY*
200 ANNOUNCEMENT: JUNIOR POSTDOCTORAL FELLOWSHIPS, CORNELL UNIVERSITY

INDEX OF AUTHORS

ARTICLES

Ariew, Roger The Duhem Thesis, 313
Ben-Zeev, Aaron The Passivity Assumption of the Sensation–Perception Distinction, 327
Butterfield, J. Relationism and Possible Worlds, 101
Currie, Gregory Individualism and Global Supervenience, 345
Erwin, Edward The Standing of Psychoanalysis, 115
Gärdenfors, Peter The Dynamics of Belief as a Basis for Logic, 1
Jennings, Richard C. Truth, Rationality and the Sociology of Science, 201
Lad, Frank The Calibration Question, 213
Price, Huw Mellor, Chance and the Single Case, 11
Schlagel, Richard H. A. Reasonable Reply to Hume's Scepticism, 359
Shanon, Benny Meno—A Cognitive Psychological View, 129
Shimony, Abner Contextual Hidden Variables Theories and Bell's Inequalities, 25
Sorensen, Roy A. Problems with Electoral Evaluations of Expert Opinions, 47
Wilkes, Kathleen V. Is Consciousness Important?, 223

DISCUSSIONS

Ehring, Douglas Probabilistic Causality and Preemption, 55
Franklin, Allan Are Paradigms Incommensurable?, 57
Glassen, Peter O'Hear on an Argument of Popper's, 375
Grenander, M. E. Behaviour Therapy: A Response to Edward Erwin, 149
Howson, Colin Bayesianism and Support by Novel Facts, 245
Lai, Tyrone The Philosophical Relevance of 'Technically Good' Experiments, 156
McCausland, Ian Einstein and Special Relativity: Who Wrote the Added Footnotes?, 60
Nissim-Sabat, Charles A Gedankenexperiment to Measure the One Way Velocity of Light, 62
O'Hear, Anthony Reply to Glassen, 377

REVIEW ARTICLES

Franklin, Allan The Epistemology of Experiment (review of Ian Hacking: *Representing and Intervening*), 381
Gibbins, Peter Nancy Cartwright's New Philosophy of Physics (review of Nancy Cartwright: *How the Laws of Physics Lie*), 390
Good, I. J. A Bayesian Approach in the Philosophy of Inference (review of Paul Horwich: *Probability and Evidence*), 161
Katouzian, Homa T. S. Kuhn, Functionalism, and Sociology of Knowledge (review of B. Barnes: *T. S. Kuhn and Social Science*), 166
Krips, Henry Popper, Propensities, and Quantum Theory (review of K. R. Popper: *Realism and the Aim of Science*, *The Open Universe*, and *Quantum Theory and the Schism in Physics*), 253
Redhead, Michael Unification in Science (review of C. F. Von Weizsäcker: *The Unity of Nature*), 274
Torretti, Roberto Space-Time Physics and the Philosophy of Science (review of Michael Friedman: *Foundations of Space-Time Theories, Relativistic Physics and the Philosophy of Science*), 280
Worrall, John An Unreal Image (review of Bas C. Van Fraassen: *The Scientific Image*), 65

REVIEWS

Apt, A. J.	D. T. Whiteside (*ed.*): *The Mathematical Papers of Isaac Newton, Volume VIII: 1697–1722*, 303
Blair, Gordon	Bryan Magee: *The Philosophy of Schopenhauer*, 179
Brown, James M.	Anthony O'Hear: *Karl Popper*, 86
D'Agostino, Fred	David Lightfoot: *The Language Lottery: Towards A Biology of Grammars*, 408
D'Agostino, Fred	Pere Julia: *Explanatory Models in Linguistics: A Behavioral Perspective*, 408
Dillinger, Mike	J. J. Katz: *Language and other abstract objects*, 301
French, Steven	P. M. Harman: *Energy, Force and Matter* and *Metaphysics and Natural Philosophy*, 297
Hart, W. D.	J. R. Maze: *The Meaning of Behaviour*, 411
Kilmister, C. W.	K. Blackwell *et al.* (*eds.*): *Cambridge Essays 1888–1899* by Bertrand Russell, 403
Lindley, D. V.	P. Gärdenfors, B. Hansson and N.-E. Sahlin (*eds.*): *Evidentiary Value: Philosophical, Judicial and Psychological Aspects of a Theory*, 293
Mellor, D. H.	Henry, E. Kyburg Jr.: *Epistemology and Inference*, 175
Milne, Peter	R. Weatherford: *Philosophical Foundations of Probability Theory*, 95
Moss, J. M. B.	D. C. Stove: *Popper and after. Four modern irrationalists*, 307
Nola, Robert	Gerard Radnitzky and Gunnar Andersson (*eds.*): *The Structure and Development of Science*, 184
Pratt, Vernon	V. J. Hensen and R. Harré (*eds.*): *The Philosophy of Evolution*, 81
Raine, Derek	N. T. Roseveare: *Mercury's Perihelion from Le Verrier to Einstein*, 188
Schaffer, Simon	Charles Webster: *From Paracelsus to Newton: magic and the making of modern science*, 191
Stevenson, Leslie	R. Nozick: *Philosophical Explanations*, 83
Waterlow, Sarah	Edward Hussey (*ed.* and *tr.*): *Aristotle's Physics Books III and IV*, 404
Weiner, Joan	D. Van Dalen (*ed.*): *Brouwer's Cambridge Lectures on Intuitionism*, 90

Brit. J. Phil. Sci. 35 (1984), 1–10 *Printed in Great Britain*

501.05
B774

The Dynamics of Belief as a Basis for Logic

by PETER GÄRDENFORS

1 *Programme*
2 *States of belief*
3 *Epistemic inputs and propositions*
4 *Basic postulates for propositions*
5 *Negation and disjunction*
6 *Completeness results*
7 *Concluding remarks*

1 PROGRAMME

The aim of this article is to give a reconstruction of propositional logic from an epistemological basis. The fundamental notions will be states of belief and changes of belief. A central idea in the reconstruction is that *propositions* are identified with a certain kind of changes of belief. On this interpretation, a proposition is characterised by the change it would induce if added to a state of belief. Formally, propositions will be defined as functions with states of belief as arguments and values. These functions will be assumed to satisfy certain requirements connected with how the addition of the information of a proposition would affect the content of a state of belief. A proposition will be said to be a tautology if it does not change any state of belief. The logic of such propositions will then be investigated. Depending on what propositions are assumed to exist one obtains different answers. It will be shown that intuitionistic propositional logic comes out very naturally on this approach and that the difference between intuitionistic and classical logic can be given a simple interpretation.

2 STATES OF BELIEF

The ontologically fundamental entities in the reconstruction of propositional logic will be *states of belief.* A state of belief is most naturally conceived of as a mental state of some person at some time. Mental states are very complex and we know very little about how beliefs are represented and

Received 1 March 1982

handled in our brains. It is not my intention to give an account of what a state of belief *is*, but, following this, I will present some different kinds of models that philosophers have used when discussing epistemological issues. In the more technical part of the paper, no assumptions will be made about the structure of the states of belief. This means that all the models of states of belief to be presented are compatible with the technical construction.

Apart from states of belief, the only entities that will be assumed are *functions* which take states of belief as arguments and values. These functions will be interpreted as representing different ways of *changing* states of belief.

The most well-known models of states of belief are the Bayesian models used in decision theory. Here a state of belief is represented by a probability measure defined over some object language or over some space of events. It is often assumed that all information about the world that is relevant for decision making is conveyed by such a probability function. This representation of states of belief is then complemented with the thesis that rational changes of belief can be described by conditionalisation of probability functions, whenever this process is applicable. The conditionalisation process can be applied only when the added information is considered certain (the posterior probability is one) and has an initial probability which is greater than zero. Jeffrey [1965] has suggested a generalisation of conditionalisation which is applicable also when the added information has a posterior probability which is less than one. This generalisation and the changes induced by other kinds of epistemic inputs to the Bayesian models are discussed in recent papers by Domotor [1980] and van Fraassen [1980].

A second kind of model which is simpler than the Bayesian, but also less informative, represents a state of belief by a set of sentences, taken from some given object language. The intended interpretation is that such a set consists of the sentences which the agent accepts in the modelled state of belief. In order to make the set of sentences more manageable, it is assumed to be closed under logical consequence. Such models are presented by Ellis [1979] and Gärdenfors [1979], [1981]. Changes of belief which can be represented by the acceptance of a new sentence can be described by adding this sentence to the earlier set of accepted sentences and then forming the deductive closure. However, it is difficult to see how other kinds of changes of belief could be modelled on this approach.

A third way of modelling states of belief is to represent a state of belief as a set of possible worlds. The interpretation of such a set is that the agent knows that the actual world is a member of the set, but it is also compatible with his beliefs that any of the worlds in the set could be the actual one. Sometimes the agent is also assumed to know the probability of each of the worlds in the set. In this case one obtains essentially a Bayesian model of the kind presented earlier. If the state of belief is expanded by the addition of the information contained in some sentence, this change can be represented by restricting the set of possible worlds to the subset where the added sentence is true. This and other forms of changes of belief modelled by possible worlds are discussed by Harper [1977], [1978].

The fourth and final type of model I wish to consider represents states of belief by computer states (in relation to some appropriate programme). In many applications within Artificial Intelligence the state of the computer is taken to be a system of beliefs about the world often described as a *semantic network*. The programme for the computer is normally constructed in such a way that by using an appropriate form of input the system of beliefs can be changed. An interesting recent example of this way of modelling states of belief is Doyle's [1979] 'truth maintenance system'. In most programmes of this kind, it is also possible to compute logical consequences of the information that is fed into the computer. The computer representation of beliefs is then used *e.g.* for planning a sequence of actions that will lead to some desired goal or to answer questions put by some user of the programme.

The four kinds of models of states of belief that have been presented here do not form a complete list. My aim is only to give some reference points for the interpretation of the formal apparatus in the following. The four kinds of models are not independent of each other. I have already mentioned a connection between the Bayesian models and the models in terms of possible worlds together with a probability measure. The relations between the Bayesian models and the models based on sets of sentences are analysed in Ellis [1979] and Gärdenfors [1979]. It is interesting to note that all four kinds of models in their standard presentation presuppose some logical concepts—the definition of a probability measure for the Bayesian models presupposes the notion of incompatible sentences, the models in terms of sets of sentences presuppose the sentential connectives and the notion of logical consequence, etc. It should be mentioned that in the following no such concepts are assumed, but they will rather be a result of the reconstruction of logic.

3 EPISTEMIC INPUTS AND PROPOSITIONS

In all the types of models presented above a state of belief changes as a result of some form of input. Depending on what type of model is chosen these 'epistemic inputs' may take different forms. The most common interpretation of such inputs is that they are the results of observations or other forms of experiences. Van Fraassen [1980] presents some types of input that are relevant for the Bayesian models of states of belief.

Here, I will avoid all problems connected with a more precise description of inputs and I will simply identify an input with the change it induces in a state of belief. Formally, this idea can be expressed by defining an epistemic input as a *function* from states of belief to states of belief. When a function representing a certain epistemic input is applied to a given state of belief K, the value of the function is the state which would be the result of accommodating the input to K. If two inputs always produce the same new state of belief, *i.e.*, if the inputs are identical as functions, there is no epistemological reason to distinguish them. This way of defining epistemic input via changes of belief is analogous to defining events via changes of physical states.

The most important type of input for the models of states of belief presented above is when new evidence is accepted as certain or 'known' in the resulting state of belief. Below, I will concentrate on this type of input and postpone the discussion of other types to the concluding section. Input corresponding to accepting evidence as certain represents the simplest kind of expanding knowledge. The information contained in such an input will be called a ***proposition***. Following the general identification of inputs presented above ***propositions are*** thus ***defined as functions from states of belief to states of belief***. This definition will be the point of departure for the reconstruction of propositional logic from the dynamics of belief.

As a first application of this definition, a central epistemological concept can now be introduced: A proposition A is said to be ***accepted as known in the state of belief*** K if and only if $A(K) = K$, or, in words, if the function corresponding to the epistemic input A is applied to K, then the resulting state of belief is K itself.

It is important to keep in mind that not all functions defined on states of belief are called propositions. Propositions correspond to a certain type of epistemic input, to wit, when new evidence is accepted as certain. In order to characterise the class of propositions, I will next formulate some postulates for propositions. Before this is done, we cannot speak of the *logic* of propositions.

4 BASIC POSTULATES FOR PROPOSITIONS

First we need a definition of the basic epistemic structure. A *belief model* is a pair $\langle \mathbf{K}, \mathbf{P} \rangle$, where $\mathbf{K}$ is a set and $\mathbf{P}$ is a class of functions from $\mathbf{K}$ to $\mathbf{K}$. Members of $\mathbf{K}$ will be called states of belief and they will be denoted K, K', ... The elements in $\mathbf{P}$ will be called propositions. A, B, C, . . . will be used as variables over $\mathbf{P}$. It should be noted that nothing is assumed about the structure of the elements in $\mathbf{K}$.

It will be assumed that the epistemic inputs corresponding to propositions can be iterated and that the composition of two such inputs is also a proposition. The composition of two propositons A and B will be denoted $A \wedge B$ (note that $A \wedge B$ does not denote a linguistic entity, but a function from states of belief to states of belief). Formally, this requirement is expressed in the following postulate:

(*P1*) For every A and B in $\mathbf{P}$, there is a function $A \wedge B$ which is also in $\mathbf{P}$ such that, for every K in $\mathbf{K}$, $A \wedge B(K) = A(B(K))$.

It follows from this postulate that the composition operation is associative, *i.e.*, that $A \wedge (B \wedge C)(K) = (A \wedge B) \wedge C(K)$, for all A, B, and C in **P** and all K in **K**.

It will also be postulated that the composition operation is commutative and idempotent:

(*P2*) For every A and B in **P** and for every K in **K**, $A \wedge B(K) = B \wedge A(K)$.
(*P3*) For every A in **P** and for every K in **K**, $A \wedge A(K) = A(K)$.

These two postulates are not valid for epistemic inputs in general. Often the order of inputs is important for which states of belief will be the result and iterations of the same input often have a cumulative effect. However, on the interpretation of propositions given above—as expansions of knowledge resulting in the acceptance of evidence as certain—these postulates are reasonable idealisations.

If A is accepted as known in K, *i.e.*, if $A(K) = K$, and $K' = B(K)$, then it follows from (*P2*) that A is accepted as known in K' too, since $A(K') = A(B(K)) = A \wedge B(K) = B \wedge A(K) = B(A(K)) = B(K) = K'$. This means that the changes of belief represented by propositions are irreversible—once a proposition is accepted as known it will be accepted in all later states of belief as long as the epistemic inputs are propositions. However, if the epistemic inputs do not satisfy the postulates for propositions, there is no guarantee that a proposition which is accepted as known will not be derogated in some later state of belief.

We can now introduce a consequence relation between propositions: A proposition B is a *consequence* of a proposition A in a belief model $\langle \mathbf{K}, \mathbf{P} \rangle$, if and only if $B(A(K)) = A(K)$, for all K in **K**. An equivalent way of formulating this is to say that B is a consequence of A, if and only if B is accepted as known whenever A is accepted as known.

For mainly technical reasons, the identity function, here denoted by T, will be assumed to be a proposition:

(*P4*) The function T, defined by $\mathrm{T}(K) = K$, for all K in **K**, is in **P**.

This function will play a central role when the logic of propositions is to be discussed. Let us say that a proposition A is a *tautology* in the belief model $\langle \mathbf{K}, \mathbf{P} \rangle$, if and only if $A(K) = \mathrm{T}(K)$, for all K in **K**. A central question is then what propositional forms are tautologies in *all* belief models. However, before a substantial answer to this question can be given, some further postulates are necessary.

The next postulate will be a formal characterisation of the information obtained when one learns that one thing *implies* another (or is *equivalent* to another). The basic idea is that if one comes to accept 'if A, then B' as certain, then this knowledge may not directly influence one's behaviour, but one will be *disposed* to accept B, if A is later accepted as certain. In order to describe this kind of belief in a formal manner, I will borrow a concept from category theory (*cf.* Goldblatt [1979], §3.10):

($P5$) For every A and B in **P**, there is a function C in **P** such that
   (a) for all K in **K**, $A(C(K)) = B(C(K))$;
   (b) for any function D in **P** such that $A(D(K)) = B(D(K))$ for all K in **K**, there is a function E in **P** such that $D(K) = E(C(K))$, for all K in **K**.

A function C which satisfies (a) and (b) will be called an *equaliser* of A and B. The interpretation of (a) is that the input corresponding to C makes the effect of the input corresponding to A equivalent to that corresponding to B. The meaning of (b) is that C corresponds to the *weakest* information in **P** with this property. (b) is sometimes also expressed by saying that any function D satisfying (a) 'factors through' C.

With the aid of ($P2$) it can be shown that there is only one equaliser of A and B in **P**. We can thus give the proposition postulated in ($P5$) a well-defined name: the equaliser of A and B will be denoted $A \leftrightarrow B$. From this we define the proposition $A \rightarrow B$ which corresponds to the information that A implies B, by the equation $(A \rightarrow B)(K) = (A \leftrightarrow (A \wedge B))(K)$, for all K in **K**.

I have now introduced enough postulates to be able to start investigating which propositions are tautologies. In order to simplify the notation in the sequel, $A = B$ will mean that $A(K) = B(K)$ for all K in **K**. This notation does not cause any confusion as long as the discussion concerns only one belief model at a time. When determining which propositional forms are tautologies in all belief models, the following lemma is extremely helpful:

Fundamental lemma: For all propositions A, B, and C in all belief models, $B \wedge A = C \wedge A$, if and only if, $(B \leftrightarrow C) \wedge A = A$.

Proof: Suppose first that $B \wedge A = C \wedge A$ in some belief model $\langle \mathbf{K}, \mathbf{P} \rangle$. This means that $B(A(K)) = C(A(K))$ for all K in **K**. By (b) in ($P5$), there exists a proposition E in **P** such that $E \wedge (B \leftrightarrow C) = A$. Then it follows that $A \wedge (B \leftrightarrow C) = E \wedge (B \leftrightarrow C) \wedge (B \leftrightarrow C)$. But according to ($P3$), $E \wedge (B \leftrightarrow C) \wedge (B \leftrightarrow C) = E \wedge (B \leftrightarrow C) = A$ and hence this half of the lemma is established.

For the converse, assume that $(B \leftrightarrow C) \wedge A = A$. From (a) in ($P5$) it follows that $B \wedge A = B \wedge (B \leftrightarrow C) \wedge A = C \wedge (B \leftrightarrow C) \wedge A = C \wedge A$. This completes the proof of the lemma.

An immediate corollary to this lemma is that, for all propositions A and B in all belief models, $A = B$ if and only if $A \leftrightarrow B = \top$. With the aid of this result it is now possible to determine some propositional forms that will be tautologies in all belief models. It can be shown that all instances of the following proposition schemas are tautologies in all models (these and the following results follow from more general results proved in Gärdenfors [1982]):

(1) $(A \wedge B) \rightarrow A$

(2) $(A \wedge B) \rightarrow B$

(3) $(A \rightarrow B) \rightarrow ((A \rightarrow C) \rightarrow (A \rightarrow (B \wedge C)))$

(4) $(A \rightarrow (B \rightarrow C)) \rightarrow ((A \rightarrow B) \rightarrow (A \rightarrow C))$

(5) $A \rightarrow (B \rightarrow A)$.

Remember that the formulas (1)–(5) are not sentences but functions on states of belief. It is also easy to show that Modus Ponens preserves tautologies, in the sense that if A and $A \rightarrow B$ are tautologies in a belief model, then B is a tautology in that model. These facts are needed when answering the question of which logic is determined by the postulates for propositions.

5 NEGATION AND DISJUNCTION

So far, the postulates that have been introduced concern only composition of propositions, corresponding to conjunction, and equalisers, corresponding to equivalence (implication). I will now turn to postulates for propositions corresponding to negation and disjunction.

The negation of a proposition will be defined by first assuming the existence of a 'falsity' proposition:

(*P6*) In every belief model $\langle \mathbf{K}, \mathbf{P} \rangle$ there exists a constant function $\perp$ in $\mathbf{P}$, *i.e.* there is some $K_\perp$ in $\mathbf{K}$ such that $\perp (K) = K_\perp$ for all K in $\mathbf{K}$.

$K_\perp$ will be called the ***absurd*** state of belief. With the aid of (*P2*) it is easy to show that $\perp$ is the only constant function in $\mathbf{P}$, since if $*$ is a function in $\mathbf{P}$ such that $*(K) = K_*$ for all K in $\mathbf{K}$, then $K_* = *(\perp(K)) = \perp(*(K)) = K_\perp$. The interpretation of the proposition $\perp$ is then given by the fact that it is accepted as known only in the absurd state of belief $K_\perp$. Furthermore, $K_\perp$ is an 'attractor' state in the sense that no further propositional input can lead out of it, since for any A, $A(K_\perp) = A(\perp(K)) = \perp(A(K)) = K_\perp$.

As is standard in propositional logic, the ***negation*** $\sim A$ of a proposition A is defined as the proposition $A \rightarrow \perp$. From (*P6*) it follows, among other things, that all propositions of the following form are tautologies in all belief models:

(6) $\perp \rightarrow A$.

The postulate for disjunction will be in the same style as the postulate concerning equalisers. When one first obtains the information that A or B, and then later comes to accept A as known, the net effect is the same as merely adding A to what one accepts (and analogously for B). This is one property that will be used to characterise the disjunction of two propositions. The other property is that the disjunction is the ***strongest*** proposition with the property mentioned above. Formally, this can be expressed as follows:

(*P7*) For every A and B in $\mathbf{P}$, there is a function C in $\mathbf{P}$ such that
(a) for all K in $\mathbf{K}$, $A(C(K)) = A(K)$ and $B(C(K)) = B(K)$;

(b) for any function D in $\mathbf{P}$ that satisfies (a), there is a function E in $\mathbf{P}$ such that $E(D(K)) = C(K)$ for all K in $\mathbf{K}$.

A function C that satisfies (a) and (b) will be called a *disjunction* of A and B. In the same way as for equalisers, it can be shown that there is only one disjunction for any pair of propositions. The disjunction of A and B will be denoted $A \vee B$.

It can now be established that all propositions of the following forms are tautologies in all belief models:

(7) $A \rightarrow (A \vee B)$

(8) $B \rightarrow (A \vee B)$

(9) $(A \rightarrow C) \rightarrow ((B \rightarrow C) \rightarrow ((A \vee B) \rightarrow C))$.

6 COMPLETENESS RESULTS

I have now introduced postulates for operations on propositions which correspond to each of the standard sentential connectives. In this section I will present some technical results which answer the question of which logic is determined by these postulates.

The formulas (1)–(9) are not sentences in a language but functions defined on states of belief. Given the postulates (*P1*)–(*P7*) there is, of course, an obvious one–one correspondence between the propositions in a belief model and the sentences in a standard 'propositional' language ($\top$ and $\bot$ are sentential constants in this language). This entails that we can consider the propositions which are tautologies in a given belief model as a class of sentences and then ask how the formulas which are included in all such classes can be axiomatised.

If interpreted as sentence schemata, the formulas (1)–(9), together with Modus Ponens as the only inference rule, constitute an axiomatisation of intuitionistic propositional logic (*cf.* Segerberg [1968]). I have shown that all instances of these schemata are tautologies in all belief models, and, as was mentioned earlier, Modus Ponens preserves tautologies. This means that the logic generated by the postulates (*P1*)–(*P7*) includes the intuitionistic propositional logic. In fact, it can be shown that the logic generated by these postulates is *exactly* the intuitionistic logic. This follows essentially from the fact that any pseudo-boolean (Heyting) algebra for intuitionistic logic can be used to construct a belief model which is equivalent to the pseudo-boolean algebra in the sense that an element is identical with the unit element in the pseudo-boolean algebra if and only if the corresponding proposition is a tautology in the belief model (this construction is presented in Gärdenfors [1982]).

It is interesting to note that if we pick out the postulates for propositions that concern '$\wedge$' and '$\rightarrow$' (or '$\leftrightarrow$') only, *i.e.*, (*P1*)–(*P5*), then it follows from the so-called separation theorem for intuitionistic logic (*cf.* Segerberg

[1968], section 7) that the logic generated by these postulates is axiomatised by the axiom schemata which contain only these connectives, *i.e.*, (1)–(5). Similarly, the logic generated by (*P1*)–(*P6*) which concern '$\wedge$', '$\rightarrow$', and '$\sim$' is axiomatised by the formulas which contain these connectives only, *i.e.*, (1)–(6); and the logic generated by (*P1*)–(*P5*) and (*P7*) ('$\wedge$', '$\rightarrow$', and '$\vee$') is axiomatised by (1)–(5) and (7)–(9), which, incidentally, is an axiomatisation of Johansson's minimal logic (*cf.* Segerberg [1968], p. 30).

In order to derive the classical propositional logic, we need one more postulate for the class of propositions in a belief model. The following additional requirement on equalisers is easy to understand given the interpretation of propositions as corresponding to the acceptance of new evidence in a state of belief:

(*P8*) For every A and B in $\mathbf{P}$, $A \leftrightarrow B = \sim A \leftrightarrow \sim B$.

Or in other words, the information which, if added to a state of belief, equalises A and B also equalises $\sim A$ and $\sim B$. It can now be shown that the logic generated by (*P1*)–(*P8*) is exactly the classical propositional logic.

7 CONCLUDING REMARKS

In this paper I have shown how propositional logic can be constructed from epistemic dynamics. The key idea for the construction is the definition of a proposition as a function representing changes of belief. This forms the conceptual basis for the postulates for propositions.

It should be noted that for each of the operations on propositions which correspond to the sentential connectives there is a postulate which only assumes the *existence* of the compound proposition, to wit, the postulates (*P1*), (*P5*), (*P6*), and (*P7*). This gives us an independent interpretation of each of the sentential connectives.

The ontological basis of the construction is very meagre. The only entities that have been assumed to exist are states of belief and functions defined on states of belief. In order to emphasise this further, let me mention some things that have *not* been assumed: Firstly, it is not necessary that there be an object language which expresses the propositions to be studied. The logic of propositions can be dealt with independently of language. Secondly, no set theory has been used, but all constructions have been expressed solely in terms of functions. Thirdly, it can be noted that the construction does not use the concept of a possible world in any way.

The fact that only some objects (called states of belief) and functions on these objects have been used reminds one of category theory. The concept of an equaliser has in fact been borrowed from that theory. Mathematically speaking, the construction presented in this paper is but a simple exercise in category theory. On the other hand, the construction uses only a very small part of the conceptual apparatus of the theory as presented by *e.g.* Goldblatt [1979].

The postulates that have been presented here are all motivated by the suggested interpretation of propositions as representing the addition of new accepted evidence to a state of belief. There are, however, other kinds of epistemic input which do not fit in with this interpretation and which do not satisfy the postulates introduced here. Van Fraassen [1980] presents some other types of input for the Bayesian model of states of belief. He also discusses the problem of how states of belief ought to change in response to these kinds of input. It would be interesting to investigate whether it is possible to formulate postulates of the same kind as those above for other kinds of epistemic input. On analogy with the construction presented here, these kinds of input would correspond to other types of functions on states of belief. It may turn out that in order to formulate such postulates, it is necessary to require that the states of belief have some additional structure, *e.g.* as is done in the Bayesian models.

An interesting structural condition on the states of belief would be to assume that they contained beliefs about the existence and identity of (physical) objects. In that case it would seem possible to postulate the existence of propositional functions corresponding to what is expressed in a quantificational language. I imagine that such an approach would help us better understand the problems connected with quantification into intensional contexts.

Lund University

REFERENCES

DOMOTOR, Z. [1980]: 'Probability kinematics and representation of belief change', *Philosophy of Science*, **47**, pp. 384–403.

DOYLE, J. [1979]: 'A truth maintenance system', *Artificial Intelligence*, **12**, pp. 231–72.

ELLIS, B. [1979]: *Rational Belief Systems*. Blackwell.

GÄRDENFORS, P. [1979]: 'Conditionals and changes of belief' in I. Niiniluoto and R. Tuomela (*eds.*): *The Logic and Epistemology of Scientific Change* (*Acta Philosophica Fennica*, **30**, issues 2–4), pp. 381–404. North-Holland.

GÄRDENFORS, P. [1981]: 'An epistemic approach to conditionals', *American Philosophical Quarterly*, **18**, pp. 203–11.

GÄRDENFORS, P. [1982]: 'Propositional logic based on the dynamics of belief', mimeographed. Department of Philosophy, Lund University.

GOLDBLATT, R. [1979]: *Topoi: The Categorial Analysis of Logic*. North-Holland.

HARPER, W. L. [1977]: 'Rational conceptual change', *PSA 1976*, vol. 2, pp. 462–94. Philosophy of Science Association.

HARPER, W. L. [1978]: 'Bayesian learning models with revision of evidence', *Philosophia*, **7**, pp. 357–67.

JEFFREY, R. C. [1965]: *The Logic of Decision*. McGraw-Hill.

SEGERBERG, K. [1968]: 'Propositional logics related to Heyting's and Johansson's', *Theoria*, **34**, pp. 26–61.

VAN FRAASSEN, B. C. [1980]: 'Rational belief and probability kinematics', *Philosophy of Science*, **47**, pp. 165—87.

Brit. J. Phil. Sci. 35 (1984), 11–23 *Printed in Great Britain*

Mellor, Chance and the Single Case

by HUW PRICE

1 *Introduction*
2 *Rationality*
3 *Chances and Colours*
4 *Chances: Redundant or Essential?*
5 *Levels of Evidence: Multiple Chances?*
6 *Are There Beliefs about Chances?*

1 INTRODUCTION

Philosophical accounts of probability differ in their sensitivity to the problems which arise from the fact that probabilities are often applied to single cases, rather than to general classes. Frequency theories of probability are often criticised for their apparent inability to make sense of such applications. One such critic is D. H. Mellor, whose own account of probability I shall be discussing here. Among recent writers on probability, Mellor perhaps pays most attention to the problems of the single case—but not with complete success, it seems to me.

It has not always been clear what constitutes a single case application of probability. In my view, sentences involving such a thing are those in which the probabilistic reference can be construed as taking the form of a sentential operator; *i.e.*, sentences of the form *Pq*, where *P* is a probabilistic operator (such as 'It is probable that . . .') and *q* a non-probabilistic sentence. Briefly, this is because a single case probabilistic judgement, *inter alia*, is one which could form the immediate basis of a choice of betting odds; and any bet is a bet *that q*, for some sentence *q*. I use the term *SP sentence* (*statement, utterance, etc.*) for a sentence (statement, utterance, *etc.*) of this form.[1]

Of central concern is thus the semantic role of a probabilistic sentential operator. Does it modify *sense*, or modify *force*? Most writers on probability assume the former answer (which accordingly I call 'the popular view').[2] Mellor himself is difficult to classify. Some recent remarks suggest he is sympathetic to the latter approach. Yet the bulk of his work on probability provides one of the more subtle accounts of the former kind. More on this anomaly in Section 6; until then, I shall be discussing Mellor's positive theory of probability, with particular attention to whether it does support the former answer to the above question; *i.e.*, as to whether it does enable *SP*

---

[1] '*SP*' for 'single case probability'.

[2] An early proponent of the 'unpopular view' was Toulmin [1950]; more recent sympathisers include Ellis [1979], Blackburn [1980] and Forrest [1981].

Received 7 August 1982

utterances to be construed as *assertions*, distinguished from assertions in general in virtue of a characteristic kind of *sense*.

It will be useful to distinguish two broad classes of popular views: those which interpret 'It is probable that . . .' along the lines of 'It is reasonable to have a strong partial belief that . . .'; and those which do *not* employ the notion of partial belief, in explicating the typical content of an *SP* sentence.[1] I shall call these the 'rationalist' and 'objectivist' approaches, respectively.

2 RATIONALITY

Though Mellor claims to make sense 'of chances being objective, empirical and not relational, and applying to the single case' ([1971], p. 175), he is in my terms more a rationalist than an objectivist. His 'main claim' is 'that chance statements assert some degrees of belief to be made more reasonable than others by objective empirical features of the world' ([1971], p. *xii*). His *propensities* are not (as in other accounts) a variety of objective probability, but rather dispositions to display chance distributions (*i.e.*, distributions of degrees of rational partial belief) over possible outcomes.

Mellor recognises the problem, common to rationalists,[2] of saying what it *means* to assert that a particular degree of partial belief is *reasonable*, in a given (single) case. His answer depends on the following argument, itself designed to show that when chance is 'specified as a fact justifying degrees of belief', it 'will relate to frequency in the way we know it does' ([1982], p. 57):

> Suppose a degree p of belief in a coin landing heads is justified by knowing that p is its chance of doing so. To get what follows from this supposition about the frequency of heads on repeated tosses, I appeal to the laws of large numbers. By these it follows in particular that as high a degree p^* of belief as we like, short of 1, is justified in any frequency proposition F of the following form: in enough such tosses, the frequency of heads would be within δ of p, where δ is any positive real number, however small.
>
> This very high degree p^* of belief in the propositions F looks as if it needs another chance to justify it; and if that were so, there would be an arguably vicious regress. But p^* does not need a chance to justify it, as the following argument shows.
>
> In Chapters 1 and 2 of (Mellor [1971]) I argued that a degree of belief increasing towards 1 must turn into a full belief before it gets there . . . For a belief to be justified, therefore, its justified degree need not be 1; it need only be a degree sufficiently close to 1. This being so, it follows that as the justified degree of a belief tends to 1, it will turn into justified belief somewhere before it gets there. Where it does so will depend on context, but that is immaterial here, since the laws of large numbers can get the justified degree of our belief in any F as close as any context could conceivably require. So whatever the context, our assumption justifies us in simply believing

---

[1] These categories are not exhaustive: there is also the *literal subjectivist* view, which takes an *SP* utterance to *report* its speaker's possession of a partial belief. But this view is easily refuted, by the observation that that it makes no sense of ordinary disputes about probabilities.

[2] Rationalists offer reductive truth-conditions for *SP* sentences, in terms of ascriptions of rationality to partial beliefs; and hence can fairly be asked to justify the claim that these rationality ascriptions are themselves truth-conditional.

> every *F*, and hence in believing that, in a sufficiently long run, the frequency of heads would come indefinitely close to the chance *p*. . . .
>
> Chances specified as facts justifying degrees of belief do therefore entail the corresponding hypothetical long run frequencies ([1982], p. 58).

Relying on this argument, Mellor offers the following account of 'what it is for a degree of belief to be justified by a fact about an event':

> The sense of justification we require has to be factual rather than, say, moral: an outcome of the coin toss is to some degree to be expected, not to be approved or deplored. Yet justification here is not a matter of making a belief true, since truth applies to the content of belief, not to its strength. There needs to be some other mode of justification by facts which relates specifically to the degree of a belief rather than its content.
>
> To see what this mode of justification is, consider . . . the connection between betting and degree of belief. . . . I argued in (Mellor [1971]) that people's choices of coherent betting quotients (CBQs) show how strong they think their beliefs are, provided they suppose the betting situation to be restricted in specified ways in order to exclude any effect on their choice of attitudes other than the belief whose strength is to be measured. I then used the entailment just established to show that only at a CBQ = *P* can I know that I would break even in a long enough run of bets on coins landing heads when their chance of doing so is always *p*. Now under the restrictions needed to make my CBQs measure my belief, . . . breaking even is the best result I could possibly hope to know of. So I have a plain gambling rationale for choosing this CBQ in that situation; and hence for having the degree of belief which, in that situation, this CBQ measures ([1982], pp. 58–9).

This account of the relevant notion of justification relies on an argument in which the notion itself plays a central role. It is not obvious that the circularity is vicious, but I think it is.

The laws of large numbers are consequences of the standard axioms of probability, and are hence satisfied in any domain which models these axioms; in particular, in the domain of coherent partial beliefs. Someone who believes to degree *p* that a given coin will land heads when tossed should, for coherence, hold an appropriately high degree of partial belief that the frequency of heads in a given large number of such tosses will be very close to *p* (if they hold any belief at all about this frequency, at least, and if they regard the tosses as independent). In other words, coherence requires that someone who holds a partial belief will, if the matter arises, adopt the corresponding effectively full belief[1] that in the sufficiently long run, the frequency of whatever it is would come indefinitely close to the degree of the partial belief. In a gambling context of the kind Mellor describes, such a person will believe that so long as they always choose the CBQ corresponding to their degree of belief, they will break even in the long run.

Mellor doesn't want to say that any partial belief is justified, so long as it coheres with the rest of a person's beliefs. Justified belief, he claims, provides not simply a full belief in breaking even in the long run, but the *knowledge* that one would do so—a *true* and *justified* effectively full belief. This claim rests on the main argument, connecting chance 'specified as a fact justifying degrees of belief' with hypothetical long run frequency. That argument runs with an as yet undefined notion of justification; and yet it is crucial to the argument that an effectively full belief which is *justified* in this undefined sense, is also justified in the sense standardly associated with full beliefs (whatever that may be), and therefore true. As we have seen, if a partial belief is said to be justified if it coheres with the rest of the holder's beliefs, the limiting argument licensed by the laws of large numbers shows that the corresponding full belief in a hypothetical long run frequency will be justified in the same sense; but this is no guarantee that it is true, or, for example, causally justified. Hence it needs to be established that the argument is dealing with a notion of justification which, so to speak, behaves properly in the limiting cases. But how is this possible, if the conclusion of this argument is required in order to define the notion of justification for a partial belief?

---

[1] The notion of a full belief is an idealisation: no one is ever so confident of a belief as to act as if it were true, no matter what the consequences of being wrong. But a belief may be ***effectively*** full, relative to some given context, in which there is a limit to the (positive and negative) utilities of the outcomes which may result from an agent's acting as if it were true.

It might be said that the relevant notion of justification is theoretical, having no reductive definition; the required limiting case behaviour being taken as axiomatic. But this leaves justification defined, effectively, in terms of the hypothetical long run frequency; or in terms of whatever it is that is held to justify (in the sense in which *justified* true belief is knowledge) a true belief with respect to the long run frequency. And these are options which Mellor explicitly rejects:

> What has a disposition to produce a long run frequency of heads on other, mostly nonexistent, tosses to do with my prospects of getting heads on this actual toss? It is no use doing what Hacking ([1965], p. 135) does with his 'frequency principle', namely in effect defining the concept of justification or support to be such that this disposition supplies it. That just provokes the question: why should I adopt for this toss the degree of belief that is justified in that sense? And to that question I know of no sufficient answer ([1982], p. 57).

In other words, if the required notion of justification is introduced as a theoretical concept, anchored by the assumption that it corresponds to the standard notion in the limiting cases, it leaves open what is then the central question: why ought we to believe what is justified, in this theoretical sense?[1]

---

[1] I have quoted here from Mellor's [1982], which is his reply to Salmon's [1979] review of Mellor's [1971]. Salmon criticises the original version of Mellor's argument; but at the wrong point, it seems to me. The problem is not whether 'any of [the] second level probabilities mentioned in the law of large numbers represents a reasonable degree of belief' (Salmon [1979], p. 192)—for that follows from the proof of the law, once it is assumed that the first level probabilities are reasonable degrees of belief (and satisfy the required axioms). It is rather that there is no non-circular way of showing that the notion of reasonableness involved behaves properly in the limiting case.

3 CHANCES AND COLOURS

Mellor draws an analogy between knowledge of chance and knowledge of colour. The above passage continues:

> We cannot get to justified degrees of belief starting from frequencies, actual or hypothetical, or from dispositions to produce them. The only way to start, as with colour, is by specifying the fact by the degree of belief it is supposed to justify. And if we do that, the frequencies, as we [have seen], . . . take care of themselves ([1982], p. 57).

Specifying single-case probabilities by the degrees of belief they justify is the essence of rationalism. As Mellor emphasises, its great advantage is to avoid the objectivists' problem with what may be called 'the downward inference'; *i.e.*, to provide a ready answer to the question: why should we adopt partial beliefs, and act, in accordance with our estimates of chance? But it lays the burden of providing truth-conditions for *SP* utterances on a notion of single-case rationality which itself stands in need of adequate explication.

The analogous step in the colour case is simple in comparison. The fact that there is a (more-or-less) invariant causal connection between an object's being of a certain colour and the resulting state of belief of a normal observer gives a relatively clear single-case sense to the notion of a colour belief being justified. The close association between 'justified' and 'true', said of full beliefs, depends on the single-case application of the notion of justification involved. I have argued that Mellor offers no adequate alternative notion of justification for the chance case.

However, it might seem plausible that the adoption of partial beliefs is also a causal matter: that standard humans react to certain kinds of situations in standard ways, by adopting certain partial beliefs. Given that standard humans are 'programmed' in a particular way, their adoption of these beliefs would then be a direct result of the relevant features of the situations concerned.

This is a promising approach, but it doesn't yield truth-conditions for *SP* utterances, or for the rationality ascriptions which the rationalist takes to paraphrase them. Such ascriptions cannot be read as reporting, or *describing*, such a causal inference. For to read 'It is reasonable to have a partial belief of degree *p* that *q*' along the lines of 'A standard human is lead to adopt a partial belief of degree *p* that *q* in these circumstances' will simply invite the question: is it reasonable to behave in the standard way in this case? In the colour case the analogous 'A standard human would see this as green' prompts an analogous question: is it *really* green? In both cases the question is neither senseless or trivial, as it would be if the state of belief resulting from the causal inference were *about* the inference itself, in the way proposed.

As it stands, only in the chance case is it likely to be seriously proposed (in an attempt to provide truth-conditions for *SP* sentences) that the beliefs and utterances which result from the causal inferences concerned, are *about* these inferences. It would be different if substantial objections were to be raised to the truth-conditionality of colour ascriptions. But colour ascriptions could not then be made assertive by being taken to be *about* the effects of situations on the minds of standard humans. Just as with chances, such a proposal would not adequately represent the ordinary use of colour ascriptions (which might therefore be taken to be non-assertive expressions of 'colour attitudes' towards objects).

4 CHANCES: REDUNDANT OR ESSENTIAL?

In my view (argued in my [1982] and [1983]) the strongest reason for denying that *SP* utterances are truth-conditional rests on a consequence of their relational character: the fact that if two speakers have different bodies of evidence, they may endorse incompatible *SP* utterances without either of them being *mistaken*. For, briefly, if two speakers express incompatible beliefs with respect to a matter of fact, then at least one of them has either inferred the belief in question from a mistaken evidential belief or perception, or has mistakenly inferred it from correct such beliefs or perceptions. *SP* utterances, in contrast, admit 'no-fault' disagreements; and are therefore not assertions as to matters of fact.[1]

The argument would fail if *SP* utterances were not relational; in particular, if they could be interpreted as implicitly *referring* to the evidence on which they appear simply to *rely*. The aim would thus be to eliminate the (apparent) relational character of such utterances, by taking them to *describe* the relationship in virtue of which they seem relational. But this proposal faces a dilemma: if the notion of *evidence* involved is taken subjectively, then although the resulting expressions are assertions, they are either the wrong assertions, or are associated with a non-standard use of the terms 'true' and 'false' in contexts in which different speakers base *SP* judgements on different bodies of evidence. (For in describing previous *SP* utterances as 'true' or 'false', one doesn't give an opinion on whether the previous speakers have correctly assessed *their* evidence, but rather indicates what follows from one's own.) If on the other hand the relevant notion of evidence is taken objectively, there will be many situations in which *SP* utterances depend on evidence of a less objective level. In this case the proposed truth-conditional account faces redundancy: for if some other account can make sense of contexts in which *SP* utterances depend on evidence less than this objective base, what is to prevent us using this other account, rather than the truth-conditional one, in cases in which the levels of evidence coincide?

Mellor appears to take the latter course, and makes it clear from the start that he is not offering an account of all *SP* contexts:

---

[1] Note that it is compatible with this—indeed a consequence of it—that when (as often happens) speakers can be assumed to be relying on the same body of evidence, their *SP* utterances will function as truth-conditional; for more details, see my [1982] and [1983].

> My project is . . . limited. I am concerned only with statistical probability, which I call 'chance'. The chances of coins landing heads, of people dying and of radioactive atoms decaying concern me; the probabilities inconclusive evidence perhaps lends to hypotheses on these and other matters do not ([1971], p. xi).

As the base of evidence on which a chance ascription depends, Mellor takes the extent of what is causally available, to a being of our limited 'conceptual and perceptual abilities' ([1982], p. 66). This is sufficiently objective to avoid the situation in which two speakers in conversation are referring to different bodies of evidence: if two speakers are in sufficient causal contact to enable them to converse, then presumably what is causally available to one is causally available to the other. Hence Mellor is able to claim that chance is non-relational. However, there are many *SP* contexts in which the evidence on which speakers rely, and on which they base some action, is less than is causally available to them. (Consider the doctor who says to her patient: 'Your operation has probably been successful. We could find out for sure, but since the tests are painful and expensive, it is best to avoid them.') If an adequate account can be given of our behaviour in these contexts—as presumably it can—why should it fail to deal with the contexts to which an account such as Mellor's applies?

Mellor seems to feel that this redundancy argument goes the wrong way: 'Inductive probabilities are, I suspect, all descended from chances', he says ([1982], p. 49); arguing as follows:

> Perhaps however there does not need to be a chance of snow, only an inductive probability of it relative to truth-making facts about temperature and pressure? Not so. Bayesian decision theory does need chances, as Jeffrey . . . has remarked. Suppose I am a smoker deciding whether to give up the habit because of my fear of cancer. For me, smoking 'dominates' not smoking, *i.e.* I shall prefer to smoke whether I have cancer or not. But I also know I should very much prefer not to have cancer, and I think cancer much more probable if I smoke. In short, the degrees of my relevant beliefs and desires make the theory tell me to give up smoking. And so I should, but only if what justifies the degrees of my conditional beliefs are propensities, not merely probabilities. Since I prefer to smoke in any case, I would be a fool to quit if my smoking were merely better evidence than my not smoking for the hypothesis that, whether I smoke or not, I shall get cancer. Quitting can only be justified if it is an action which will cause a change in my prospects, namely a reduction in the probability of my getting the disease. But a probability which has causes is a part of the physical world, not merely part of inductive logic; that is . . . a chance. In other words, inductive probabilities are not enough to make sense of . . . Bayesian decision theory. The work objective probability has to do there can only be done by chance ([1982], pp. 62–3).

I don't want to defend inductive probabilities as a means of supplying truth-conditions to *SP* utterances. But I do dispute that objective chances are essential to decisions of the smoking kind. Consider the following case: I have been arrested for robbery, in circumstances which leave me no defence. I know I will be convicted, but would much prefer a suspended sentence to a spell inside. A sympathetic court official, in charge of scheduling cases, offers me a choice of Judge *A* or Judge *B*. Judge *A*, he tells me, gives a suspended sentence (in the event of conviction, and where possible) in every second case on his list; Judge *B*, in every third case on hers. He explains that although he knows where I will be on the list of whichever judge I choose —and hence whether I will get a suspended sentence—it is more than his job's worth to tell me that much. So I have to choose without this extra information. And I should choose Judge *A*: even if I should slightly prefer, *ceteris paribus*, to appear before Judge *B* than Judge *A*, so that for me Judge *B* 'dominates' Judge *A*; my utility table being as follows:

	Suspended sentence	Spell inside
Judge *A*	10	−10
Judge *B*	12	−8

Whichever judge I choose, however, it is only a matter of objective chance whether I get a suspended sentence in the trivial sense that my chance of doing so is 0 or 1—if 'objective chance' is taken in Mellor's sense, at any rate. On the evidence causally available it is effectively certain whether or not I get a suspended sentence, once I have chosen my judge; and there is no guarantee that in choosing Judge *A* I increase my chance of a suspended sentence—I may make it 0.

The example rests on the fact that not all the evidence causally available to both the court official and me is in fact available to me, given the practical constraints of the situation. So it might be objected that chance ascriptions refer implicitly not to the evidence causally available *in principle*, but to what is available *in practice*. But this is too subjective: it allows users of conflicting *SP* utterances to be referring to different bodies of evidence. I have already mentioned the problems this leads to for the view that *SP* utterances are assertions.

Alternatively, it might be said that in such a case objective chances are a useful fiction, though recognized not to be 'real'. However, in the supposedly analogous cases on which this claim would depend—cases in which scientific theories are held to be used as useful fictions—it is important that there be some explanation of the fact that a theory which is acknowledged to be false is nonetheless useful. Such explanations are of several kinds. Sometimes the false theory provides a close approximation in a certain range of cases, even though it has been replaced in general by a theory of wider scope; then the new theory justifies the continued use of the old one. Or a theory may depend on assumptions which are known from the start to be false, but which are regarded as useful idealisations (*e.g.*, the assumption of continuity in the treatment of fluids). The fact that such an assumption does not lead to serious error seems usually to be explicable in terms of the theory in virtue of which the assumption is regarded as false (in the fluid case, in terms of molecular theory, and particularly the size of molecules in relation to ordinary fluid bodies).

No account of these kinds seems possible in the chance case. There is no relevant sense in which my choice of Judge *A* approximates to one based on a real objective chance (in Mellor's sense). And since my position on the court list of whichever judge I choose determines whether my chance of receiving a suspended sentence is 0 or 1, the objective chance account does not justify my acting in ignorance of this position.

Perhaps my decision approximates that of someone faced with the same choice, but whose position on the court list is not causally available to them, even in principle. Assuming that such a person has no greater chance of receiving one position than another on the list of whichever judge they choose, their objective chance of a suspended sentence is 1/2 if they choose Judge *A*, and 1/3 if they choose Judge *B*. My decision resembles this person's, the claim would be, so I should act in the same way.

However, such an account is in great danger of being *too* successful. If it establishes unreservedly that it is reasonable for us to behave as someone should for whom the evidence available in principle, is just the evidence we have in fact, it will run into an objection raised by Ayer to certain types of relational and frequency accounts of probability: that it is unclear why, given the choice, we should prefer probability judgements based on larger amounts of evidence.[1] No matter how little evidence we have, we can imagine someone to whom no more is available, even in principle. If it is therefore reasonable to act in the way that this person should, according to what are for him the objective chances, then why should we ever bother to obtain more evidence? More general second-level evidence might indicate that the more first-level evidence we pay attention to, the more chance we have of making successful decisions. But why should this concern us, if we have yet to be convinced of the advantages of taking into account more than our present (first-level) evidence?

So it seems that objective chances, in Mellor's sense, are no help in explaining and justifying my choice of Judge *A* rather than Judge *B*. Yet this simple piece of Bayesian decision theory differs significantly from Mellor's smoking example only in the nature of the obstacle which prevents the agent concerned obtaining more information relevant to his decision. The smoker knows that smokers encounter carcinogens more frequently than do nonsmokers, and wishes to avoid such contact himself. But he doesn't know whether if he continued to smoke he would actually encounter such substances in sufficient quantity to give him cancer (or whether he would do so—or has already done so—even if he now gives up smoking). The prisoner knows that people tried by Judge *B* receive gaol sentences more frequently than those tried by Judge *A*, and wants to avoid one himself. But he doesn't know whether he would in fact do so even if he went before Judge *B*, or fail to do so before Judge *A*. (It would be easy to add to the present example something analogous to individual variation in susceptibility to a given carcinogen.) There seems to be no difference between the two cases sufficient to show that objective chances play an essential role in guiding and justifying our behaviour in the former one, given that they play no such role in the latter.

---

[1] See Ayer [1957], and [1963], ch. 7.

5 LEVELS OF EVIDENCE: MULTIPLE CHANCES?

Even if the redundancy argument of the last section is put aside, there are still problems with the class of contexts an account such as Mellor's will take to involve objective chances. For one thing, contexts in which a group of speakers share a common base of evidence, with the result that their *SP* utterances appear truth-conditional,[1] differ a great deal with respect to the level of objectivity of this evidence. A group of doctors discussing a patient's chances of survival may all be relying on the same evidence; but know that if another piece of equipment was available, say, their estimate might be quite different. The same patient's friends might share another body of evidence, and dispute his or her chances on that basis. In general when a given event or proposition instantiates various different descriptions or generalisations, different theories may ascribe it different chances. In the context of any one such theory the relevant *SP* expressions will function truth-conditionally, but their doing so does not provide evidence of a real objective chance. For which theory-relative chance would it be?

In other words, the objectivism of usage which results from the fact that a group of speakers may base their *SP* utterances about a certain matter on the same body of evidence, does not indicate that such speakers are referring to non-relational objective chances, as Mellor proposes. Many such contexts actually preclude this interpretation.

This kind of problem is not peculiar to Mellor's view of chance, but plagues all the standard accounts. A hypothetical long-run frequency view, for example, relies on the idea of the repetition of the situation to whose outcome a chance is to be applied. But as is often pointed out, what counts as a repetition depends on how the situation is characterised. Saying what is to be included in the characterisation is just the problem of specifying the relevant level of evidence. However, Mellor has an unusual solution: he allows for 'an event instantiating more than one statistical law, attaching different chances to the same result'. 'There is nothing paradoxical in one outcome of an event having two chances', he says, 'even if it is uncommon. All it means is that two degrees of the same belief could both be objectively justified; and that can be so even if no one could actually have them both at once' ([1982], pp. 66–7).

The problem with this is Ayer's objection, noted above. If an event has different chances under different descriptions, in what sense can it be better to adopt a partial belief corresponding to one rather than another? And what motive can there be for preferring a chance based on more evidence to one based on less? It is not clear that there isn't a satisfactory answer; but it seems there can be none in terms not available to a frequentist, say, for whom the parallel question is, 'Why should we act on the basis of the most complete available frequency evidence?' Mellor claims to do better than frequentists, in avoiding this kind of question. But he can only do so if chances are unique and non-relational—a state of affairs which we now see receives less support from the apparent truth-conditionality of much of the ordinary use of *SP* expressions than might at first be supposed.

---

[1] See note 1, p. 16.

6 ARE THERE BELIEFS ABOUT CHANCES?

I want to finish with some remarks about a respect in which Mellor's account seems less in accord with the popular view than are most accounts of single-case probability. Mellor says that the reason 'it has proved so hard to frame an acceptable account of objective chance' is that 'people naturally feel that, if chance is objective, it must make true beliefs with some characteristic content', 'supposing that the only objectifying job facts can have to do is making beliefs true' ([1982], p. 61).

> But this is not so, as we may see in ethics. An objectivist there will naturally give truth-conditions, such as: '*X* is good' is true if and only if *X* is good. This truth-condition, like that for chance, is not trite; to give it, the objectivist has to think there are facts such as *X* being good, and that *X*'s having some other properties would make the fact obtain. But he is not thereby committed to thinking that this fact's role is to make a true belief whose content is that *X* is good. Obviously not: its role is objectively to justify a mental state quite different from belief, namely approving of *X*. The content of all the relevant *beliefs* about *X* is entirely non-moral. Morality, like chance, has no subject matter of its own. It may be objective nonetheless; beliefs are not the only mental states capable of objective justification. Failure to see this underlies two views in ethics which correspond closely to frequency and subjective views of chance. One tries to provide a distinctive content for the belief that *X* is good; e.g. that *X* promotes human happiness. The other sees that this misses the whole point of morality (since it remains an open question whether one should approve of promoting human happiness or anything else), and concludes that there is no objective goodness at all.
>
> Properly to recognise both objective goodness and objective chance, we need to extend our conception of facts as suppliers only of true beliefs ([1982], pp. 61–2).

Note, first, that the views to which Mellor is objecting here can agree that the *ultimate* role of a chance fact is to justify a partial belief. Objectivist views which recognise the question 'Why does a full belief about a chance make reasonable a corresponding partial belief?' (as all such views should) do so because they recognise that the key role of chance facts is in some sense to guide our partial beliefs and hence our actions. The issue here is thus whether there are really full beliefs about chance facts, as a step in the process whereby this function is carried out. Mellor is apparently denying that there are such beliefs. I think he is right, but it is a view which seems to me to be difficult to reconcile with other aspects of his account.

One problem is this: given that we do say things like: 'There is a chance of

Q 16285

1/2 that this coin will land heads when I next toss it', Mellor apparently has to claim either (*i*) that such an utterance does not make an assertion about one of the objective chances whose existence he accepts; or (*ii*) that it does make such an assertion, but that what it asserts cannot be believed to be the case in the usual way. If (*i*), then what could be the evidence for thinking that there are objective chances (given that apparent assertions about such things are in fact to be taken as examples of some kind of non-assertive idiom)? But if (*ii*), then what notion of belief could be involved? And what notion of assertion? On the usual view, an assertion characteristically results from, and is a key display of, a corresponding effectively full belief. Moreover, if there are objective chances then someone who is prepared to make an assertion such as the above one will presumably exhibit the non-linguistic behaviour which would ordinarily be taken as displaying the disposition in which the full belief associated with this assertion would consist. That is, roughly, such a person will behave as if there is a chance of 1/2 that the coin will land heads; they will be willing to bet a large stake for a small return that there is this chance, for example, and in general will behave in whatever way is to their advantage if there is such a chance. What non-arbitrary grounds—other than an acceptance that there are really no such things as objective chances—could there be for denying that they actually have the belief of which this behaviour is the display, according to the usual picture of belief? Of course we would expect such a person also to have the disposition characteristic of a partial belief of degree 1/2 in a result of heads. Saying *why* they should hold the latter disposition, given that they hold the former one, is for an objectivist the task of justifying the downward rule of inference. It doesn't help in doing so to refuse to call the former disposition a belief.

Mellor says that 'for me to think there is a chance of heads is . . . for me to have a degree of belief in heads and to think it justified . . . by some fact about the coin toss' ([1982], p. 60). So to *say* there is a chance of heads is presumably to express this partial belief and to say, at least implicitly, that it is justified. It is remarks like these which give Mellor's account its rationalist character. Rationalism demonstrates that so long as there is an adequate notion of single-case rationality, or justification, *SP* utterances and beliefs—including those which mention chances—can quite well be construed as assertions and beliefs about rational partial belief. Mellor thinks that there is such a notion of justification (we have seen that he takes pains to define it), and at some points seems close to the rationalist use of it. So from his point of view his rejection of beliefs about chances 'with some characteristic content' seems unnecessary. So although I am in sympathy with the move itself, I think it fits uneasily with other features of Mellor's account.[1]

The Australian National University, Canberra

---

[1] This paper gives me the opportunity to acknowledge my debt to Hugh Mellor for help and encouragement in philosophy over a number of years. I am grateful to him for enabling me to run the debt up—and for the rate of interest he has managed to maintain.

REFERENCES

AYER, A. J. [1957]: 'The Conception of Probability as a Logical Relation', in S. Korner (*ed.*), *Observation and Interpretation*. London: Butterworths, pp. 12–17; [1963]: *The Concept of a Person*. Macmillan.

BLACKBURN, S. [1980]: 'Opinions and Chances', in D. H. Mellor (*ed.*), *Prospects for Pragmatism*. Cambridge University Press, pp. 175–96.

ELLIS, B. [1979]: *Rational Belief Systems*. Oxford: Blackwell.

FORREST, P. [1981]: 'Probabilistic Modal Inferences', *Australasian Journal of Philosophy*, **59**, pp. 38–53.

HACKING, I. [1965]: *Logic of Statistical Inference*. Cambridge University Press.

MELLOR, D. H. [1971]: *The Matter of Chance*. Cambridge University Press; [1982]: 'Chances and Degrees of Belief', in R. B. McLaughlin (*ed.*), *What, Why, Where, When*. Reidel, pp. 49–68.

PRICE, H. [1982]: 'Does "Probably" Modify Sense?', forthcoming; [1983]: '"Could a Question be True?": Assent and the Basis of Meaning', forthcoming in *The Philosophical Quarterly*.

SALMON, W. C. [1979]: 'Propensities: a Discussion Review of D. H. Mellor, *The Matter of Chance*', *Erkenntnis*, **14**, pp. 183–215.

Brit. J. Phil. Sci. 35 (1984), 25–45 *Printed in Great Britain*

Contextual Hidden Variables Theories and Bell's Inequalities

by ABNER SHIMONY

Noncontextual hidden variables theories, assigning simultaneous values to all quantum mechanical observables, are inconsistent by theorems of Gleason and others. These theorems do not exclude contextual hidden variables theories, in which a complete state assigns values to physical quantities only relative to contexts. However, any contextual theory obeying a certain factorisability condition implies one of Bell's Inequalities, thereby precluding complete agreement with quantum mechanical predictions. The present paper distinguishes two kinds of contextual theories, 'algebraic' and 'environmental', and investigates when factorisability is reasonable. Some statements by Fine about the philosophical significance of Bell's Inequalities are then assessed.

1 *Introduction*
2 *Noncontextual Hidden Variables Theories*
3 *Contextual Hidden Variables Theories*
4 *Bell's Theorem*
5 *Philosophical Significance*
6 *Comments on Fine's Papers*
7 *Conclusions*

1 INTRODUCTION

In most of the discussions of Bell's Inequalities little notice is taken of the historical circumstances of their discovery. J. S. Bell was told by J. Jauch of the work of A. Gleason [1957] which showed that simultaneous values cannot be assigned to all observables of a quantum mechanical system in a way that respects their algebraic structure (except in the simple case in which the relevant Hilbert space has dimension less than three). Gleason's theorem thus precludes a type of hidden variables theory that has come to be called 'noncontextual'. The question of agreement between the statistical predictions of quantum mechanics and those of a noncontextual hidden variables theory does not even arise, since such a theory cannot be consistently formulated. Bell noticed [1966], however, that there is a large family of hidden variables theories, which have come to be called 'contextual', that are not precluded by Gleason's and related theorems. He found, in fact, that the hidden variables theories which had been developed in the literature in some detail were contextual, even though their authors had not articulated the distinction. Bell noticed further that one of these

Received 22 September 1982

theories, that of Bohm [1952], postulated a peculiar kind of nonlocality or action at a distance in order to recover the quantum mechanical predictions concerning correlated, spatially separated systems of the kind studied by Einstein, Podolsky, and Rosen [1935]. Bell asked whether such nonlocality is a necessary condition for the recovery of the statistical predictions of quantum mechanics. His positive answer to this question [1964] is called 'Bell's Theorem'. (The paper of 1966 was written earlier than that of 1964 but because of an editorial mishap was published later.) Bell proved the theorem by showing that a certain factorizability condition upon a hidden variables theory implies that certain statistical correlations between two spatially separated systems always satisfy an inequality which he formulated, while quantum mechanics implies that this inequality is sometimes violated. The factorisability condition is reasonable when the two systems are separated by typical laboratory distances (a few metres), and especially reasonable if the operations performed upon the two systems are events with space-like separation. Bell concluded that no physically acceptable contextual hidden variables theory could agree completely with the statistical predictions of quantum mechanics. Since the experiments inspired by his work overwhelmingly support quantum mechanics, it follows that no contextual hidden variables theory is viable unless it postulates a kind of nonlocality. A survey article by J. Clauser and A. Shimony [1978] gives details of the history of this topic up to its publication date, but it must be supplemented by the very important paper of Aspect, Dalibard, and Roger [1982].

One of the purposes of this paper has been achieved in the first paragraph, by pointing out that Bell's Theorem was the result of a question concerning the physical adequacy of contextual hidden variables theories. A further purpose is to examine the concept of contextual hidden variables theory in more detail than is common in the literature. It is a complex concept, and at least two very different types can be discriminated—to be called 'algebraic' and 'environmental'. This conceptual study will help to clarify the content and the philosophical significance of Bell's Theorem and the Inequalities involved in it.

In two recent papers A. Fine [1982a, b] presented some theorems concerning the relation between Bell's Inequalities and joint probability distributions, and on the basis of these theorems he drew some conclusions about the philosophical significance of the Inequalities. The theorems are correct and interesting, but I wish to challenge the statements about philosophical significance. Specifically, I believe that these statements neglect the importance of contextual hidden variables theories.

2 NONCONTEXTUAL HIDDEN VARIABLES THEORIES

The set $\mathscr{L}$ of 'eventualities' or 'propositions' concerning a physical system S is a convenient entity to study, because its mathematical structure is considerably simpler than that of the set of observables of S. I prefer the term 'eventuality', which was introduced by H. Stein [1972], because it does not invite the conflation of questions about the structure of $\mathscr{L}$ with questions of deductive logic. Each eventuality is a bivalent possibility—either true or false—if it has a definite value, but the value may be indefinite in certain states of S. $\mathscr{L}$ is assumed to be partially ordered with respect to a relation $\leqslant$, where $a \leqslant b$ intuitively means that whenever a is true, b also is. There is an orthocomplementation operation $'$ on $\mathscr{L}$, such that $(a')' = a$ and $a \leqslant b$ implies $b' \leqslant a'$. The intuitive interpretation is that a' is false if a is true, and a' is true if a is false, though the possibility is open that neither has a definite truth value. In the standard Hilbert space formulation of quantum mechanics $\mathscr{L}$ is assumed to be isomorphic to the lattice $\mathscr{L}_H$ of projectors (or equivalently of closed linear subspaces) of an appropriate Hilbert space $\mathscr{H}$ (*cf.* Mackey [1963]). Unless otherwise noted, this assumption will be made throughout the present paper. This isomorphism guarantees that the g.l.b. $\wedge a_i$ and the l.u.b. $\vee a_i$ $(i = 1, 2, \ldots)$ with respect to the relation $\leqslant$ are well-defined, though of course the isomorphism to $\mathscr{L}_H$ is not a necessary condition for these σ-lattice properties to hold.

There is a standard procedure for defining observables in terms of the lattice $\mathscr{L}_H$, and a theorem of J. von Neumann establishes an isomorphism between the algebra of observables and the algebra of self-adjoint operators on $\mathscr{H}$ (see Varadarajan [1968], pp. 143–4). The eventualities can, in fact, be taken as a subclass of observables—*viz.*, bivalent observables with (conventionally) $+1$ representing truth and 0 representing falsity. The self-adjoint operators corresponding to the eventualities in accordance with von Neumann's theorem are the projectors, which of course have no other eigenvalues than 1 and 0. If e is an eventuality, E_e will denote the projector corresponding to it. For the purposes of this paper there will be no need to discuss further any observables other than eventualities.

G. Mackey [1963] introduced the important concept of a measure on $\mathscr{L}_H$, that is, a mapping $p : \mathscr{L}_H \rightarrow [0, 1]$ such that (*i*) $p(1) = 1$, where 1 is the identity operator (and hence the projector that has the whole of $\mathscr{H}$ as its range), and (ii) if the a_i are such that $a_i \leqslant a'_j$ whenever i and j are distinct $(i, j = 1, 2, \ldots)$, then

$$p(\vee a_i) = \Sigma p(a_i).$$

Clearly, Mackey's concept is similar in some ways to that of a classical probability measure, except that it is adapted to the lattice structure of $\mathscr{L}_H$, which in general is not Boolean (as the lattice of subsets of a phase space is). Quantum mechanics supplies some exemplary measures on $\mathscr{L}_H$. Let $\psi \in \mathscr{H}$ and have norm 1. If p_ψ is defined for all $e \in \mathscr{L}_H$ by the equation

$$p_\psi(e) = (\psi, E_e\psi),$$

it can easily be shown to satisfy conditions (*i*) and (*ii*) of Mackey's definition.

Convex combinations of denumerably many measures of the form p_ψ also satisfy these conditions.

Gleason proved the fundamental theorem that if the dimension of $\mathscr{H}$ is greater than 2 and p is a measure, then there is a finite or denumerable sequence of normalized vectors ψ_i and of non-negative real numbers w_i summing to 1, such that for all $e \in \mathscr{L}_H$

$$p(e) = \Sigma w_i p_{\psi_i}(e).$$

A corollary is the nonexistence of a p which assigns only 1 or 0 to each $e \in \mathscr{L}_H$ (sometimes called a 'dispersion-free measure'). The corollary was proved directly, without proceeding via Gleason's theorem, by Bell [1966], S. Kochen and E. Specker [1967], and F. Belinfante [1973]. These proofs show in fact that there are finite sublattices of $\mathscr{L}_H$ on which dispersion-free measures cannot exist.

Gleason's theorem and its corollary doomed the program of noncontextual hidden variables theories. This program envisaged a space Λ of complete states of S, each member λ of which determines a dispersion-free measure on $\mathscr{L}_H$. It envisaged, furthermore, a σ-algebra $\Sigma(\Lambda)$ of subsets of Λ and classical probability functions ρ such that each triple $\langle \Lambda, \Sigma(\Lambda), \rho \rangle$ constitutes a classical probability space. And finally it envisaged that for each normalized $\psi \in \mathscr{H}$ there is a classical probability measure function ρ_ψ on $\Sigma(\Lambda)$ such that for all $e \in \mathscr{L}_H$

$$p_\psi(e) = \int_\Lambda p_\lambda(e) d\rho_\psi. \tag{1}$$

This program was actually achieved (in the sense of a mathematical model) by Bell [1966], Kochen and Specker [1967], and Pearle [1965], when the dimension of $\mathscr{H}$ is 2. It fails when the dimension of $\mathscr{H}$ is greater than 2 because of Gleason's theorem and its corollary. The failure, of course, occurs in the first step of the program, for there are no dispersion-free states p_λ, hence no space Λ of complete states having the desired properties; and therefore one cannot even pose the question of whether there is a ρ_ψ over $\Sigma(\Lambda)$ satisfying Eq. (1).

3 CONTEXTUAL HIDDEN VARIABLES THEORIES

Contextual hidden variables theories are not excluded by these negative results. This kind of hidden variables theory was exemplified by the models of L. de Broglie [1926], D. Bohm [1952], and N. Wiener and A. Siegel [1953], but the significance of the difference from noncontextual theories was first enunciated by Bell [1966]. Again a space Λ of complete states and a σ-algebra $\Sigma(\Lambda)$ of subsets of Λ are envisaged. Now, however, p_λ is conceived to be a function of two variables e and C, where e is a bivalent possibility (perhaps an eventuality of S, but other options will later be noted), and C is the 'context' in which the value of e is determined. Bell defends the idea of contextual hidden variables theories by saying, 'The result of an observation may reasonably depend not only on the state of the system (including hidden variables) but also on the disposition of the apparatus' [1966, p. 451]. The character of the contexts, however, is not the same in all expositions.

Gudder [1970] developed in considerable mathematical detail a contextual hidden variables theory in which the context C is taken to be some subset of the lattice $\mathscr{L}_H$. It is reasonable to call this type of theory 'algebraic'. Specifically, Gudder proposes to let C be a maximal Boolean sub-σ-algebra of $\mathscr{L}_H$, and $p_\lambda(e, C)$ is well defined if and only if $e \in C$. When $\mathscr{H}$ has dimension greater than 2, an eventuality e can belong to more than one maximal Boolean sub-σ-algebra, and therefore the specification of C is not redundant. It is assumed that $p_\lambda(e, C)$ is either 1 or 0 (hence the measure is dispersion-free), and Mackey's conditions (*i*) and (*ii*) are satisfied within each C. If e belongs both to C and C', the possibility is left open that $p_\lambda(e, C) \neq p_\lambda(e, C')$. Not only do Gleason's theorem and its corollary not preclude the existence of such m_λ, but the remainder of the hidden variables program can be carried out: for each normalised $\psi \in \mathscr{H}$ there is a classical probability measure ρ_ψ such that

$$p_\psi(e) = \int_\Lambda p_\lambda(e, C)\, \mathrm{d}\rho_\psi. \tag{1'}$$

Hence Gudder has provided a mathematical existence proof of a contextual hidden variables theory which agrees with the statistical predictions of quantum mechanics, whatever the quantum state of S may be.

Physically, however, Gudder's proposal is not fully satisfactory, since rarely does an actual apparatus single out a maximal Boolean sub-σ-algebra along with the eventualities which are to be observed. If ideal apparatus should do so, then Eq. (1′) assures agreement with quantum mechanics, but it is not obvious that agreement would be achieved when the apparatus is non-ideal. The same difficulty would arise concerning other contextual hidden variables theories of the algebraic type.

It is desirable, therefore, to consider another type of contextual hidden variables theory, which may appropriately be called 'environmental', in which the context C is a state of the physical environment (including apparatus) with which the system S interacts. (Heywood and Redhead [1983] also use the term 'environmental', but in a somewhat different sense.) In principle C could be the complete state of the environment. Then according to a deterministic theory C together with the complete state λ of the system S determine the value of any relevant bivalent possibility e. In practice, of course, the environment is very imprecisely known and characterized. Consequently, there is motivation to let C be a 'coarse-grained state' of the environment, such as is specified by a macroscopic description of the apparatus. It is reasonable to assume that the complete state of S, together with a coarse-grained state of the environment, suffice to determine a probability $p_\lambda\ (e, C)$ that the bivalent possibility e will have a designated value (say $+1$). Good experimental design in microphysical investigations is, in fact, intended to establish situations in which such probabilities are well defined: situations which exhibit sensitivity to the microscopic features of the system S of interest, sensitivity also to the macroscopic features of the apparatus. There is no *a priori* guarantee that a hidden variables theory will achieve these two types of sensitivity, so that $p_\lambda(e, C)$ will be well defined. One that does not do so, however, does not seem to be capable of making statistical predictions about the outcomes of actual experiments and therefore would be of little interest to working scientists.

4 BELL'S THEOREM

The question may now be raised: can a contextual hidden variables theory of the environmental type account for all the experimental data that quantum mechanics explains? The question is not sharply formulated, because the class of theories is ill-defined. I suppose, however, that the question could be sharpened in a natural way, and the answer would then be positive: the environmental theory would be linked to an algebraic theory, and the agreement of the latter with quantum mechanics could be established along the lines indicated by Bell and Gudder. Bell, rather elliptically, suggests such a procedure at the end of his article of 1966. He does not pursue details on this point, since a question of principle intrudes. He asks effectively (not using this language) whether a contextual hidden variables theory of the environmental type is able to recover the statistical predictions of quantum mechanics without introducing physically undesirable features. His first inequality [1964] and other inequalities inspired by his work (which I shall call collectively 'Bell's Inequalities') provide a strong negative answer to this question of principle. This result is especially remarkable in view of the looseness of the characterisation of theories of environmental type, and it is possible only because the inequalities can be demonstrated with great generality. The negative answer removes most of the physical motivation to characterise this type of hidden variables theory with precision and to investigate them in detail.

The inequality which is most convenient for the purposes of this paper is that of Clauser and Horne [1974], but I shall state the conditions under which it holds in a new way in order to emphasise generality.

Theorem: Let a be a parameter which can take on the values a', a'', and b be a parameter which can take on the values b', b''. Let $\langle\Lambda, \Sigma(\Lambda), \rho\rangle$ be a probability space: *i.e.*, Λ is a nonempty set, $\Sigma(\Lambda)$ is a σ-algebra of subsets of Λ, and ρ is a probability measure on $\Sigma(\Lambda)$, such that for each value of a and b there are random variables $p^1_\lambda(a)$, $p^2_\lambda(b)$, and $p_\lambda(a, b)$ with values between 0 and 1 for each $\lambda \in \Lambda$. Suppose that

$$p_\lambda(a, b) = p^1_\lambda(a)p^2_\lambda(b). \tag{2}$$

Then

$$-1 \leqslant p(a', b') + p(a', b'') + p(a'', b') - p(a'', b'') - p^1(a') - p^2(b') \leqslant 0, \tag{3}$$

where

$$p^1(a) = \int_\Lambda p^1_\lambda(a)\,\mathrm{d}\rho, \tag{4a}$$

$$p^2(b) = \int_\Lambda p^2_\lambda(b)\,\mathrm{d}\rho, \tag{4b}$$

$$p(a, b) = \int_\Lambda p^2(a, b)\,\mathrm{d}\rho. \tag{4c}$$

Comments: (*i*) Inequality (3) will be called 'the Bell/CH Inequality'. The proof of the theorem is given by Clauser and Horne.

(*ii*) The statement given of the theorem is purely mathematical, but it is obviously amenable to physical interpretation. Specifically, the set Λ can be interpreted as the space of complete states of a system S. The parameters a and b can label bivalent possibilities A_a and B_b (with values 1 and 0, for example), and $p^1_\lambda(a)$ and $p^2_\lambda(b)$ are respectively the probabilities that A_a and B_b will have the value 1 if the complete state is λ, while $p_\lambda(a, b)$ is the probability that A_a and B_b both will be 1 if the complete state is λ. ρ is the probability distribution over $\Sigma(\Lambda)$ associated with some procedure for preparing an ensemble of replicas of S.

(*iii*) There remain various options for interpreting A_a and B_b. (a) One obvious option is to interpret these as eventualities of S: *e.g.*, if S is a photon propagating in the $\hat{z}$-direction, A_a can be the eventuality that S is linearly polarized along a line that makes the angle a relative to a fixed $\hat{x}$-direction, and B_b has a similar meaning for a photon propagating in the $-\hat{z}$ direction. $p^1_\lambda(a) = 1$ means that A_a is true when λ is the complete state of S, $p^2_\lambda(b) = 1$ means that B_b is true in this state, and $p_\lambda(a, b) = 1$ means that the conjunction of A_a and B_b is true in this state; in each instance a value 0 means falsity. If $p^1_\lambda(a)$ and $p^2_\lambda(b)$ can only take on the values 1 and 0, and do so in a way that is compatible with Mackey's conditions (*i*) and (*ii*) on measures, then one has a fragment of a noncontextual hidden variables theory—not the whole of such a theory unless every eventuality in $\mathscr{L}_H$ is identifiable with one of the A_a and B_b. (b) Another option is to let a be an ordered pair (e, C), where C is a maximal Boolean sub-σ-algebra of $\mathscr{L}_H$ and $e \in C$. If $p^1_\lambda(a)$, etc. can assume only the values 1 and 0 and satisfy Mackey's conditions (*i*) and (*ii*) within each C, then we have a fragment of a contextual hidden variables theory of algebraic type. (c) A third option is that A_a and B_b refer to bivalent possibilities under prescribed experimental conditions. For instance, there could be an experimental arrangement in which two photons, propagating respectively in the $\hat{z}$ and $-\hat{z}$ directions, impinge perpendicularly upon sheets of polaroid, which are oriented with transmission axes making arbitrary angles to the fixed $\hat{x}$-axis. A_a can be interpreted as the bivalent possibility that the first photon will be transmitted if the first sheet of polaroid is oriented at the angle a, with $+1$ signifying transmission and 0 signifying non-transmission; and B_b has an analogous interpretation. Less elliptically, specifying the angle a is tantamount to a choice of a pair (e, C), where e is a bivalent possibility and C is a state (normally macroscopically specified) of part of the environment; and similarly with b. Since the type of environment and type of binary possibility are antecedently fixed in discussions of correlation experiments, the only variables needed to complete the macroscopic description of the experiment are the angles of orientation of the analysers. We have here a fragment of a contextual hidden variables theory of environmental type. Since the contexts C are not normally complete states of the environment, one does not expect $p^1_\lambda(a)$, $p^2_\lambda(b)$, and $p_\lambda(a, b)$ to be restricted to 1 and 0, but only to have values lying in the interval $[0, 1]$.

(iv) Eq. (2) is referred to by Clauser and Horne, following Bell, as the 'locality condition'. In the interest of generality, however, it is desirable to avoid a locution which suggests the locality of relativity theory, and therefore I shall follow Garuccio and Selleri [1978] and Fine [1982a, b] in calling it the 'factorisability condition'. Whether it is a reasonable condition to impose depends upon the options concerning A_a and B_b, and also upon our knowledge of the physical world. There are two very different cases in which it is clearly a reasonable condition. (a) In a noncontextual hidden variables theory all eventualities are assigned definite truth values in a complete state, and all observables possess definite values simultaneously. If both A_a and B_b are true, so is their conjunction; and then the interpretation of $p^1_\lambda(a)$, $p^2_\lambda(b)$, and $p_\lambda(a, b)$ automatically ensures that the factorisability condition is satisfied. This argument does not depend upon the nature of S (*e.g.*, whether or not S consists of two spatially separated parts S_1 and S_2), or on the character of the eventualities A_a and B_b (whether the first refers to part S_1 and the second to part S_2, whether the projectors corresponding to A_a for different values of a are commuting, etc.). The fact that the factorisability condition is satisfied in so general a way in a noncontextual hidden variables theory provides special motivation for avoiding the weighted term 'locality condition'. (b) Under the right conditions Eq. (2) will hold for a contextual hidden variables theory of environmental type. The conditions are the ones commonly stated in deriving Bell's Inequalities. The system S consists of two spatially separated parts S_1 and S_2. The relevant parts of the environment are two bivalent analysers, each with a channel labelled $+$ and a channel labelled $-$, one analyser for S_1 and the other for S_2. Each analyser has an adjustable feature (typically an orientation) parametrised by a and b respectively. The analysers are well separated on a laboratory scale (typically by several meters), with precautions taken that the choice of parameter of each analyser not influence the analysis performed by the other analyser. Let $p^1_\lambda(a)$ be the probability that S_1 goes into the $+$ channel of its analyser if the parameter has value a and the complete state of S is λ; let $p_\lambda^2(b)$ be the probability that S_2 will go into the + channel of its analyser if the parameter has value b and the complete state of S is λ; and let $p_\lambda(a, b)$ be the probability that both S_1 and S_2 will go into their respective + channels, given a, b, and λ. Under these conditions it is reasonable that $p_\lambda(a)$ be independent of b, as the notation indicates, and that $p_\lambda(b)$ be independent of a. Also, since λ is the complete state of $S_1 + S_2$ all correlations between the behaviour of S_1 and that of S_2 should be carried by λ. If λ were not given, then the passage of S_1 into its + channel would convey some (usually probabilistic) information relevant to the behaviour of S_2, and conversely, but this is not the case if λ is given and if there is no direct causal connection between the two parts of the experiment. It follows that the factorisability condition, Eq. (2), is reasonable. Of course, some physical influence could be exerted by the choice of b upon the behaviour of S_1 and by the choice of a upon the behaviour of S_2, in spite of precautions to the contrary. In order to preclude unnoticed causal influence, it suffices (provided that relativity theory is true) for the compound event which consists of the choice of a and the passage of S_1 into a channel to have space-like separation from the corresponding compound event concerning S_2. The terminology 'locality condition' for Eq. (2) is certainly appropriate when this additional precaution about space-like separation is taken; but I think that in a commonsense way it is even appropriate without this additional precaution, provided that care is taken to prevent manifest causal influence between the two sides of the experiment. For a discussion of these points see the 1981 publications of Fine and Shimony.

The term 'Bell's Theorem' means more than Bell's Inequalities. It also comprises the fact that there are some situations in which the quantum mechanical predictions violate Ineq. (3) (and other Bell's Inequalities), so that hidden variables theories satisfying the factorisability condition cannot completely agree with the statistical predictions of quantum mechanics. This disaccord is, of course, very interesting as it stands. But of even greater interest is the fact that there are situations in which the experimental evidence violates Ineq. (3) or others of Bell's Inequalities. The experimental situation is chosen to be one in which the quantum mechanical prediction violates the Inequalities, for it is anticipated that these predictions will be correct. That expectation has been overwhelmingly fulfilled. The most impressive experiment is that of Aspect, Dalibard, and Roger [1982], which obtained results violating one of Bell's Inequalities, but agreeing with quantum mechanics, when the two compound selection-analysis events have space-like separation.

5 PHILOSOPHICAL SIGNIFICANCE

What can one conclude, with all due regard for fallibility?

First, no noncontextual hidden variables theory can be completely correct. This is so, because the factorisability condition, Eq. (2), is automatically satisfied in the noncontextual case.

Second, no contextual hidden variables theory of the environmental type can be completely correct, unless Eq. (2) is abandoned. And the experimental situation of Aspect *et al.* does not permit the abandonment of Eq. (2) without a radical change of world view. Whether the requisite change is a revision of relativistic space-time structure is too large a question to be discussed here. I have elsewhere [1978] suggested the possibility of 'peaceful coexistence' between the quantum mechanical nonlocality exhibited in the correlation experiments and relativistic space-time structure, and hope to explore the issue further. Even 'peaceful coexistence', however, requires a radical change of our conception of the way in which events occur in the theatre of space-time.

I say nothing about contextual hidden variables theories of the algebraic type, just because their connection with experiment is difficult to discuss at the level of abstraction on which they have been proposed. Silence on this point does not mean, however, that they offer a viable loophole from the results of the correlation experiments. If detailed connection is established between a theory of algebraic type and experiments, that theory will be subject to disconfirmation by the correlation experiments unless it violates the factorisability condition.

Since the predictions of quantum mechanics have been confirmed and those of its most promising alternatives have been disconfirmed, we have stronger reasons than ever for believing that the following peculiar implications of quantum mechanics will remain permanently as established parts of physical theory: (1) In any state of a physical system S there are some eventualities which have indefinite truth values. (2) If an operation is performed which forces an eventuality with indefinite truth value to achieve definiteness (whether this happens only by measurement or also by some other means is one of the outstanding problems of the foundations of quantum mechanics) the outcome is a matter of chance. (3) There are 'entangled systems' (in Schrödinger's phrase) which have the property that they constitute a composite system in a pure state, while neither of them separately is in a pure state.

If a noncontextual hidden variables theory had been viable, then some aspects of the world view of classical physics would have achieved a new lease of life: definiteness of truth values of all eventualities, elimination of chance at the fundamental level, no entanglement.

If a contextual hidden variables theory of environmental type had been viable, then a world view definitely different from that of quantum mechanics, and yet different in some respects from that of classical physics would have been suggested. In this world view a system S has a definite complete state, and so does its environment. The outcome of an experiment performed upon S is determined cooperatively by these two complete states. Hence, this kind of theory does not have the fundamental stochastic character of quantum mechanics. (A further generalisation, which would permit the outcome to be a matter of chance even when the state of the entire universe is given, will not be considered here, since our concern is with the possibility of world-views which differ sharply from that of quantum mechanics.) The entanglement of systems can be avoided if the complete state of each subsystem of a composite system is definite even when the quantum state (which from the standpoint of a hidden variables theory contains only partial information) cannot be factorised into definite quantum states for each subsystem. Instead of entanglement, which is a kinematical feature of a composite system in quantum mechanics, there would be causal interaction between components, which is a dynamical matter. The outcome of an experiment is a cooperative effect, just as it is conceived to be in classical physics. Classically, however, the system S possesses a definite array of properties, and the interaction of S with measuring apparatus produces an effect which is indexical of one or more of these properties—perfectly so if the apparatus is ideal, imperfectly with actual apparatus. A contextual hidden variables theory of environmental type differs from classical physics by regarding as relational many features which in classical physics, and also in noncontextual hidden variables theories, are thought to be intrinsic to S. Of course, exactly which features are fixed by the complete state λ of S, and which are fixed by λ together with the complete state of the environment, cannot be told without giving the details of the hidden variables theory. The contextual hidden variables theories which are most relevant to the correlation experiments performed so far are those which conceive the components of linear polarisation of photons, or the components of spin of certain atoms, to depend both upon the state of the particle involved and upon the state of the environment, and not upon the former by itself.

The philosophical significance of Bell's Inequalities, in my opinion, is that they permit a near decisive test of those world views which are contrary to that of quantum mechanics. Bell's work made possible, therefore, some near decisive results in experimental metaphysics. These results certainly do not imply that quantum mechanics as it is now formulated will never be displaced or ameliorated by any physical theory. After all, only a handful of quantum mechanical predictions are tested in the correlation experiments. This handful should not be under-estimated, however, since they concern salient, sensitive points—points at which a clash of world view occurs. The confirmation of the quantum mechanical predictions at these points and the disconfirmation of one or another of Bell's Inequalities provide strong grounds for confidence about future physical theories: any theory which displaces but improves upon quantum theory will preserve the indefiniteness of eventualities, the fundamental role of chance, and the entanglement of systems.

The whole vast body of experimental results which have probed and confirmed quantum mechanics can in fact be regarded as contributions to experimental metaphysics, and the mathematical analysis of the structure of quantum mechanics by Gleason, Kochen and Specker, and others has helped to make explicit its metaphysical content. Notably, the structure of the lattice of eventualities for all but the simplest quantum systems (all for which the appropriate Hilbert space has dimension greater than 2) precludes a noncontextual hidden variables theory, hence the possession of definite values simultaneously by all observables. This is a profound metaphysical result, which antedated Bell's Inequalities. It must be emphasised, however, that the experimental evidence for the structure of the lattice of eventualities of physical systems is fragmentary and for the most part indirect. The only case in which there is abundant direct evidence is that of photon polarisation, and this is unsatisfactory for the purpose at hand, since the quantum mechanical treatment of polarisation of a single photon requires only a two-dimensional Hilbert space. A systematic set of experiments on spin-1 atoms would suffice to establish the structure of the lattice of eventualities in the simplest relevant case, but these experiments have never been performed. Stern-Gerlach methods would suffice to establish the structure of a certain interesting sublattice of eventualities for the spin-1 system (Hultgren and Shimony [1976]), but even these have not been carried out; and the structure of the full lattice of eventualities for the spin-1 system requires methods that are much more elaborate than those of Stern-Gerlach (Swift and Wright [1980]). Consequently, a die-hard defender of noncontextual hidden variables theories might express scepticism about the premiss of Gleason and of Kochen and Specker, etc., that the lattice of eventualities is isomorphic to the lattice of projectors on a Hilbert space. This (rather perverse) defence of noncontextual hidden variables theories is defeated by the experimental violation of Bell's Inequalities. Many fewer measurements are needed for checking one of the Inequalities than for establishing the structure of the lattice of eventualities, and therefore Bell's result is a very helpful instrument even for ends which in principle could be achieved in other ways. And for the purpose of disconfirming contextual hidden variables theories which conform to relativistic restrictions upon causality, we have at present no other means than Bell's Inequalities. (A large family of local contextual hidden variables theories could in principle be refuted by using the theorem in Section 3 of Heywood and Redhead [1983] or a related unpublished result of Kochen. In order to use their results for this purpose it would be necessary to exhibit certain correlations between a pair of systems, each of which is known experimentally to have a lattice of eventualities $\mathscr{L}_H$ with $\dim \mathscr{H} > 2$. Such an experiment would be much more intricate than those based upon Bell's Theorem.)

6 COMMENTS ON FINE'S PAPERS

Fine [1982a,b] proves several new theorems concerning hidden variables, joint probability, and Bell's Inequalities, and on the basis of these theorems he makes some strong philosophical claims, which I wish to assess.

In order to state Fine's main theorem, the notation introduced in Section 4 must be extended in a natural way: $p^1(\bar{a})$ is the probability (when a specified method of preparing the system of interest is given) that the bivalent possibility A_a has the value 0, and similarly for $p^2(\bar{b})$; $p(a, \bar{b})$ is the probability that A_a has value 1 and B_b has value 0, and similarly for $p(\bar{a}, b)$ and $p(\bar{a}, \bar{b})$. The parameter a can take on values a' and a'' (and perhaps others not of immediate interest), while b can take on values b' and b'' (and perhaps others). Fine inquires about the possibility of extending the set of single and double probabilities $p^1(a)$, $p^1(\bar{a})$, $p^2(b)$, $p^2(\bar{b})$, $p(a, b)$, etc. to a set of compatible quadruple probabilities, where 'compatibility' has the natural meaning:

$$p(a', b') = p(a', a'', b', b'') + p(a', \mathrm{a}'', \mathrm{b}', \bar{b}'') + p(a', \bar{a}'', b', b'') + p(a', \bar{a}'', b', \bar{b}''), \quad (5)$$

and similarly for $p(a', b'')$, $p(a', \bar{b}')$, etc. He proves that a necessary and sufficient condition for this extension to be possible is that the singles and doubles satisfy the Bell/CH Inequality(3). This theorem is proved both in his [1982a] and his [1982b], and I shall refer to it as Fine's Theorem 3, which is its number in the latter, more comprehensive paper.

Theorem 3 is the central theorem upon which he bases the following conclusions about the philosophical significance of Bell's Inequalities: 'our investigations suggest that what the different hidden variables programs have in common, and the common source of their difficulties, is the provision of joint distributions in those cases where quantum mechanics denies them' [1982b, p. 1309], and '...hidden variables and the Bell inequalities are all about...imposing requirements to make well defined precisely those probability distributions for noncommuting observables whose rejection is the very essence of quantum mechanics' [1982a, p. 294].

I am very sceptical of Fine's strong claims. He has, in my opinion, neglected the diversity of hidden variables theories and the diversity of the aims of different programmes. I shall try to indicate briefly that the results of research on hidden variables theories are too intricate to be characterized accurately by his condensed statements. Furthermore, I doubt that his own theorems, though correct mathematically, help much to elucidate the philosophical significance of Bell's Theorem.

One class of hidden variables programmes starts with the premiss that quantum mechanics gives a correct characterisation of physical systems. Since there are various inequivalent formulations of quantum mechanics, this premiss can be interpreted in many ways, but for the present purposes we can restrict attention to Mackey's axiomatization [1963]. Upon the assumption that quantum mechanics is correct one can either have a programme of showing that the quantum mechanical formalism is susceptible of a hidden variables interpretation or a programme of demonstrating the contrary, and each of these programmes ramifies according to the meaning of 'hidden variables interpretation'. If, in particular, 'hidden variable interpretation' is understood in the noncontextual sense, then the main result was already mentioned in Section 2: no noncontextual hidden variables interpretation is possible if the dimension of $\mathscr{H}$ is greater than 2, because no dispersion-free measure $p_\lambda(e)$ can be imposed upon the lattice $\mathscr{L}_H$. If a dispersion-free measure did exist, then it would be straightforward to define the joint probability of an arbitrary set of eventualities $e_1, \ldots, e_n$, corresponding to projectors which may not pairwise commute: namely

$$\begin{aligned} p_\lambda(e_1, \ldots, e_n) &= 1 \quad \text{if} \quad p_\lambda(e_i) = 1 \text{ for each } i = 1, \ldots, n, \\ &= 0 \quad \text{otherwise.} \end{aligned} \tag{6}$$

If, furthermore, the values of the parameter λ constitute a space Λ with a σ-algebra of subsets $\Sigma(\Lambda)$, then other joint probability distributions could be obtained by integrating $p_\lambda(e_1, \ldots, e_n)$ with arbitrary probability measures over $\Sigma(\Lambda)$. The non-existence of even a single dispersion-free measure p_λ (proved by Gleason [1957], Bell [1966], Kochen and Specker [1967], and Belinfante [1973]) is thus an obstacle to the construction of joint probability distributions for arbitrary eventualities. Fine's conclusion, quoted above, about 'probability distributions for noncommuting observables whose rejection is the very essence of quantum mechanics' applies well enough to the non-existence results of Gleason *et al.*, without pretending to provide an exact summary of their results.

When the dimension of $\mathscr{H}$ is 2, Fine's conclusion about probability distributions for noncommuting observables does not apply without serious qualification. In this case, dispersion-free states p_λ can be constructed on $\mathscr{L}_H$, and Λ and $\Sigma(\Lambda)$ can be constructed so that for every quantum state ψ there is a probability measure ρ_ψ over $\Sigma(\Lambda)$ such that for every $e \in \mathscr{L}_H$

$$p_\psi(e) = \int_\Lambda p_\lambda(e)\, dp_\psi.$$

When $p_\psi(e_1, \ldots, e_n)$ is defined as in Eq. (6), then

$$p_\psi(e_1, \ldots, e_n) = \int_\Lambda p_\lambda(e_1, \ldots, e_n)\, d\rho_\psi \tag{8}$$

is a joint probability of $e_1, \ldots, e_n$, whether or not the corresponding projectors commute, and the usual $p_\psi(e_i)$ can be recovered from Eq. (8) by marginalisation. At first glance, the case of dim $\mathscr{H} = 2$ appears to provide a counter-example to Theorem 7 of Fine [1982b], namely: observables $A_1, \ldots, A_n$ satisfy the joint distribution condition if and only if all pairs commute. A counter-example does not materialise, however, because Fine uses 'joint distribution condition' in a strong sense, *viz.*, 'corresponding to every n-place Borel function f, there is an observable of the system with operator $f(A_1, \ldots, A_n)$, and corresponding to every state Ψ of the system there is probability measure $\mu_{\Psi, A_1, \ldots, A_n}$ on the Borel set of R^n that returns the quantum single distributions P_{A_i} as marginals, such that

$$\mu_{\Psi, A_1, \ldots, A_n}(f^{-1}(S)) = P_{f(A_1, \ldots, A_n)}(S)$$

for every state Ψ and Borel set S of reals'. The demand in this definition for an observable with operator $f(A_1, \ldots, A_n)$ cannot be met, even when a joint probability distribution is defined by Eq. (8). This fact can be illustrated simply by letting $n = 2$ and A_1, A_2 be noncommuting projectors representing eventualities e_1, e_2. There exists no self-adjoint operator A such that

$$p_\psi(e_1, e_2) = P^\psi_A(\{1\}),$$

where the expression on the right means the quantum mechanical probability, given ψ, that A will have its value in the (unit) set $\{1\}$; specifically, A cannot be chosen to be A_1A_2 or A_2A_1, because neither of these operators is self-adjoint. Although Fine's Theorem 7 is correct, Eq. (8) provides a reasonable way to introduce joint probability distributions in the case $\dim \mathscr{H} = 2$, by a procedure less restrictive than his. Another way of putting the matter is that if $\dim \mathscr{H} = 2$ then the lattice $\mathscr{L}_H$ can be embedded in a Boolean lattice $\mathscr{B}$ (see Kochen and Specker 1967). Not surprisingly, the greatest lower bound (glb) of A_1 and A_2 is not the same in $\mathscr{B}$ as in $\mathscr{L}_H$. In $\mathscr{B}$ the glb is an element which intuitively is the ordinary conjunction of the eventualities represented by A_1 and A_2, whereas in $\mathscr{L}_H$ the glb is the null operator, which represents the impossible eventuality. The embeddability of $\mathscr{L}_H$ in $\mathscr{B}$ underlies the existence of joint probability distributions for arbitrary eventualities in the case of $\dim \mathscr{H} = 2$.

Something remains to be said about joint distributions of arbitrary eventualities even in the case of $\dim \mathscr{H} > 2$, in spite of the non-embeddability of $\mathscr{L}_H$ in a Boolean lattice and the nonexistence of a dispersion-free measure, and hence the non-applicability of Eq. (8). Suppose that we focus upon a fragment $\mathscr{F} \subset \mathscr{L}_H$, not necessarily a sublattice, which is singled out for some reason such as the design of an experiment, and suppose also that is embeddable in a Boolean lattice $\bar{\mathscr{F}}$. Then probability distributions can be constructed on $\bar{\mathscr{F}}$, which automatically provide joint probability distributions for any set $e_1, \ldots, e_n$ of eventualities in $\mathscr{F}$. One may inquire, however, whether a probability distribution on $\bar{\mathscr{F}}$ exists which is compatible with a *pre-assigned* distribution on $\mathscr{F}$. Fine's theorem 3 gives a mathematically pleasing answer to this question in a special case. Let $\mathscr{F}$ consist of the eight eventualities a', a'', b', b'', $a' \wedge b', a' \wedge b'', a'' \wedge b', a'' \wedge b''$, and $\bar{\mathscr{F}}$ be the Boolean lattice constructed from a', a'', b', b'' by conjunctions and negations. Fine's Theorem 3 then says that a necessary and sufficient condition for the existence of probability distribution on $\bar{\mathscr{F}}$ compatible with a preassigned distribution p on $\mathscr{F}$ is that p satisfies the Bell/CH Inequality. Garg and Mermin have derived results applying to other fragments $\mathscr{F} \subset \mathscr{L}_H$, but the question concerning an arbitrary fragment still remains open. It should be mentioned, as a qualification of Fine's philosophical conclusion quoted earlier, that his Theorem 3 does not yield a straightforward connection between commutativity and the existence of joint distributions. If a', a'', b', b'' are eventualities representing by commuting projectors, then indeed for any $\psi \in \mathscr{H}$ the probability measure p_ψ satisfies the Bell/CH Inequality, and therefore joint distributions compatible with p_ψ are guaranteed. If, however, the projectors representing a' and a'' or those representing b' and b'' fail to commute, then p_ψ may or may not satisfy the Bell/CH Inequality, for the quantum mechanical probabilities violate the Bell/CH Inequality only for special cases.

The philosophical significance of Fine's Theorem 3 can be assessed by considering both cases in which the experimental data confirm the Bell/CH Inequality and those in which they disconfirm it. (*i*) If the data confirm the Inequality, then Fine's Theorem 3 permits the fragment $\mathscr{F}$ to be interpreted within a fragmentary noncontextual hidden variables theory, which, as we have seen earlier, retains in broad outlines the world view of classical physics. This is a new result, which does not seem to be accessible without Fine's theorem. However, this retention of a classical world view is only local, for the theorems of Gleason *et al.* show that definite truth values cannot be assigned to all eventualities of $\mathscr{L}_H$ if dim $\mathscr{H} > 2$, and hence a global classical world view is impossible if the formalism of quantum mechanics is correct. (*ii*) If the experimental data disconfirm the Bell/CH Inequality then not even for the fragment $\mathscr{F}$ can a noncontextual hidden variables theory be consistently formulated. This result does not require Fine's Theorem, for it is just below the surface in the earlier literature on Bell's Theorem: a noncontextual hidden variables theory automatically guarantees the factorisability condition (as pointed out in Section 4), from which the Bell/CH Inequality follows.

Let us now consider programmes which keep the premiss that quantum mechanics is correct, but admit noncontextual hidden variables theories. Such theories, in contrast to noncontextual hidden variables theories, do not explicitly aim at joint distributions for non-commuting observables. This contrast is particularly clear for contextual theories of algebraic type, but it also holds for those of environmental type. If e_1 and e_2 are eventualities represented quantum mechanically by non-commuting projectors, then there does not exist a single Boolean sub-σ-algebra to which both e_1 and e_2 belong. Hence, even though a complete state λ of S assigns a definite truth value to e_1 when a context (maximal Boolean sub-σ-algebra) to which it belongs is singled out, and likewise λ assigns a definite truth value to e_2 when a context to which it belongs is singled out, there is no one context in which definite truth values are assigned to both e_1 and e_2. Therefore, it does not even make sense to pose the question: what is the probability that e_1 and e_2 are jointly true? Of course the same strictures apply when more than two eventualities $e_1, \ldots, e_n$ not all belonging to a single maximal Boolean sub-

σ-algebra are considered. If the contextual hidden variables theory is of environmental type, then e_1 and e_2 must be interpreted as bivalent physical possibilities, and C_1 and C_2 as states of the respective environments in which these possibilities are actualized. If C_1 and C_2 are mutually exclusive (*e.g.*, they describe different orientations of a polaroid sheet), then it makes no sense to speak of e_1 and e_2 jointly having the value 1, or jointly being true, and *ipso facto* no sense to speak of the probability of this joint outcome.

Indeed, the restrictions on joint distributions imposed by contextual hidden variables theories are very much like those of quantum mechanics itself—and exactly the same in the case of Gudder's version of a contextual theory of algebraic type. This concordance is not surprising when one recalls that Bell, by a judo-like manoeuvre, cited Bohr in order to vindicate a family of hidden variables theories in which the values of observables depend not only upon the state of the system but also upon the context [1966, p. 452]. The concordance is so great, in fact, that one may wonder whether there is really any difference between contextual hidden variables theories and quantum mechanics as it is ordinarily interpreted. This question has already been answered at the end of Section 5, but it is worthwhile reiterating the answer. A contextual hidden variables theory envisages a space Λ of complete states, such that the specification of $\lambda \in \Lambda$ together with the complete context (algebraic or environmental) uniquely determines the values of all bivalent physical possibilities that are compatible with that context. Hence chance, which is a fundamental element of quantum mechanics, is denied in a contextual hidden variables theory (at least in the deterministic case to which we have restricted our attention). Probability may enter a contextual hidden variables theory, but only in the way that it enters classical statistical mechanics: via probability distributions over $\Sigma(\Lambda)$. Whether this probability should be interpreted as personal, or logical, or propensity probability, or in some other way, is not important at this point. What is important is that the deterministic contextual hidden variables theories do not admit chance as a fundamental level, but claim rather that the state of the universe (system plus total environment) suffices to determine all outcomes. Also, the contextual hidden variables theories do not admit the entanglement of systems at the level of complete states, and interpret quantum mechanical entanglement only epistemically.

It has been suggested by Fine (private correspondence) that joint probability distributions of non-commuting observables can be introduced in contextual hidden variables theories by speaking counter-factually. Suppose, as considered in Section 4, that a' is the ordered pair (e', C'), where C' is a maximal Boolean sub-σ-algebra containing e', and a'' is the analogous ordered pair (e'', C''). Then $a'(\lambda) = 1$ means that if the complete state of S is λ and the context C' is realized, then e' is true; and $a''(\lambda) = 1$ has an analogous meaning. Then $(a' \wedge a'')(\lambda) = 1$ could be interpreted as saying that if C' were the context then e' would be true, and if C'' were the context e'' would be true, all upon supposition that the complete state of S is λ. The lattice of propositions introduced in this way, which may be called $\mathscr{L}_C$, is Boolean and admits joint distributions determined by all probability measures ρ over $\Sigma(\Lambda)$. Of course, the price paid for introducing joint distributions in this way is that the new lattice $\mathscr{L}_C$ has an algebraic structure which disagrees globally with $\mathscr{L}_H$, though it agrees locally, *i.e.*, on the Boolean sublattices of $\mathscr{L}_H$. The Boolean structure of $\mathscr{L}_C$ suffices to guarantee the existence of joint distributions without any consideration of the Bell/CH Inequality or Fine's Theorem 3.

Finally, we consider hidden variables programmes which make no commitment to the quantum mechanical formalism, but inquire whether the experimental evidence can be accounted for by a contextual hidden variables theory of environmental type. Much of what was said above about joint probability distributions in contextual theories of algebraic type applies also to those of environmental type. Two different analyses can be envisaged concerning the joint probabilities of bivalent possibilities which are associated with mutually exclusive environmental contexts. One analysis denies that joint probability makes sense when there is no single experimental arrangement (or other environmental context) in which the bivalent possibilities all have definite truth values. The other analysis prescribes speaking counterfactually, so that the conjunction of a' and a'' would be well-defined even if a' and a'' refer to mutually exclusive contests. The conjunction of a' and a'' consists in asserting that e' would be true were context C' realized and also that e'' would be true if context C'' were realized. When conjunction is thus defined it makes sense to talk about the probability that the conjunction is true, and of course this analysis can be extended to quadruples a', a'', b', b'', etc. There is some good sense in each of these two analyses, and I do not wish to adjudicate between them in this paper, because to do so would require an extended investigation of counterfactuals.

Without adjudication, however, one can see how each of the two analyses would assess the applicability of Fine's Theorem 3. (*i*) From the standpoint of the first analysis, nothing of interest is lost if quadruple probabilities $p(a', a'', b', b'')$, etc. are undefinable when the Bell/CH Inequality is false, since the conjunction of a', a'', b', b'' is considered to be meaningless. On the other hand, if the Bell/CH Inequality is experimentally true, so that the quadruple probabilities exist according to Fine's Theorem 3, then the first analysis would merely say that this is a formal result without physical significance. The only significant probabilities, according to the first analysis, are those of a', a'', b', and b'' singly, and of the doubles $a' \wedge b'$, $a' \wedge b''$, $a'' \wedge b'$, and $a'' \wedge b''$, since each of these unequivocally refers to a single realisable context. (*ii*) According to the second analysis the joint probability of a' and a'', of b' and b'', and of the quadruples a', a'', b', b'', etc. are always defined, via counterfactuals, regardless of the mutual exclusiveness of the contexts. But now Fine's Theorem 3, which says that a necessary and sufficient condition for the existence of compatible quadruples is the satisfaction of the Bell/CH Inequality, becomes irrelevant. The quadruple probabilities exist whether or not the Inequality holds. The reason for this irrelevance is that the existence of compatible quadruple probabilities, when conjunctions are understood counterfactually, does not imply the factorisability condition.

Bell's Inequalities remain unequivocally relevant to the question of whether the empirical evidence can be accounted for by a contextual hidden variables theory of environmental type. Whenever the experimental arrangement and the theoretical presuppositions (notably relativity theory) imply the factorizability condition (or an appropriate variant thereof), then Bell's Inequalities provide the crucial link between theory and experiment. Specifically, in Section 4 it was argued that the factorisability condition is reasonable in the experimental arrangement of Aspect, Dalibard, and Roger, if relativity theory is assumed. However, Fine [1981] presented a case against the factorisability condition even under these circumstances. His papers [1982a, b] can be read as supplying a new argument against factorisability, even though he is not explicit on this point. He could (and perhaps implicitly does) argue as follows: Factorisability seems harmless and reasonable in the situation constructed by Aspect *et al.* But we know that the Bell/CH Inequality implies the existence of joint probabilities of a', a'', b', b'', etc. Hence, implicit in the factorisability condition is an extremely strong and far from harmless assumption about the existence of joint probabilities.

According to the first of the analyses considered above, the quadruple probabilities which are inferred from factorisability have no physical significance, and their existence is just as harmless as their nonexistence would be. According to the second analysis (which Fine endorses in correspondence), the existence of quadruple probabilities is not surprising after all. We can indeed ask about the probability that both e' would be true if its context C' were realised and e'' would be true if its context C'' were realised, and similarly for other pairs, triples, and quadruples. The existence of joint probabilities in this sense does not depend upon the truth of the factorisability condition, but only upon having a contextual hidden variables theory which is employed counterfactually. One must be very careful, however, to specify the environmental contexts accurately. In the polarisation correlation experiment of Aspect *et al.* and others, a typical context C is a state of a part of the environment of a pair of photons, specifying at least the angles of orientation of both polarisation analysers and perhaps other things as well. The Bell/CH Inequality does not follow from the existence of joint (double, triple, quadruple) probabilities when the contexts are understood in this way. An extra premiss is needed: namely, a context sufficient for fixing the probability of a specifies only the orientation of the first polarisation analyser, so that the orientation of the second polarisation analyser is irrelevant, and likewise for the probability of b. Whether this additional premiss is reasonable from the standpoint of relativistic non-quantum physics can and should be debated (as in the 1981 papers of Fine and Shimony), but this question is untouched by considerations of the existence of joint probabilities in a contextual hidden variables theory.

7 CONCLUSIONS

Cataloguing possible world views is a formidable enterprise, and cataloguing those which are compatible with a certain body of experimental evidence is yet more formidable. Some of the world views which have been proposed and discussed in the literature on the foundations of quantum mechanics have not even been mentioned in this paper. Among those which have been considered, however, the evidence has narrowed the choices—though again due regard must be given to the possibility of experimental error. The world view associated with noncontextual hidden variables theories is overwhelmingly disconfirmed, whether or not one pays attention to relativistic locality. That associated with contextual hidden variables theories is overwhelmingly disconfirmed if relativistic locality is assumed. Bell's Theorem is a useful instrument for strengthening the first of these judgments, and at present it is indispensable for achieving the latter. That, in my opinion, is the philosophical significance of Bell's Inequalities and Theorem.

Acknowledgments: I wish to thank Arthur Fine and Michael Redhead for comments which were valuable in revising this paper.

Boston University

REFERENCES

ASPECT, A., DALIBARD, J., and ROGER, G. [1982]: 'Experimental test of Bell's Inequalities using variable analyzers', *Physical Review Letters*, **49**, 1804–1807.

BELINFANTE, F. [1973]: *A Survey of Hidden-Variables Theories*. Pergamon Press.

BELL, J. S. [1964]: 'On the Einstein Podolsky Rosen Paradox', *Physics*, **1**, pp. 195–200.

BELL, J. S. [1966]: 'On the problem of hidden variables in quantum mechanics', *Review of Modern Physics*, **38**, pp. 447–52.

BOHM, D. [1952]: 'A suggested interpretation of the quantum theory in terms of "hidden variables"', *Physical Review*, **85**, pp. 166–79, pp. 180–93.

BROGLIE, L. DE [1926]: 'Sur la possibilité de relier les phénomenes d'interférence et de diffraction à la théorie des quanta de lumière', *Comptes Rendus*, **183**, pp. 447–8.

CLAUSER, J. F. AND HORNE, M. A. [1974]: 'Experimental consequences of objective local theories' *Physical Review D*, **10**, pp. 526–35.

CLAUSER, J. F. AND SHIMONY, A. [1978]: 'Bell's theorem: experimental tests and implications', *Reports on Progress in Physics*, **41**, pp. 1881–927.

EINSTEIN, A., PODOLSKY, B., AND ROSEN, N. [1935]: 'Can quantum-mechanical description of physical reality be considered complete?', *Physical Review*, **47**, *pp.* 777–80.

FINE, A. [1981], 'Correlations and physical locality', *PSA 1980*, vol. 2, ed. A. Asquith and R. Giere (Philosophy of Science Association, E. Lansing, Michigan), pp. 535–62.

FINE, A. [1982a]: 'Hidden variables, joint probability, and the Bell Inequalities', *Physical Review Letters*, **48**, pp. 291–5.

FINE, A. [1982b]: 'Joint distributions, quantum correlations, and commuting observables', *Journal of Mathematical Physics*, **23**, pp. 1306–10.

GARG, A. AND MERMIN, D. [1982]: 'Comment on "Hidden variables, joint probability, and the Bell Inequalities" ', *Physical Review Letters*, **49**, p. 242.

GARUCCIO, A. AND SELLERI, F. [1978]: 'On the equivalence of deterministic and probabilistic local theories', *Lettere al Nuovo Cimento*, **23**, pp. 555–8.

GLEASON, A. [1957]: 'Measures on the closed subspaces of a Hilbert space', *Journal of Mathematics and Mechanics*, **6**, pp. 885–93.

GUDDER, S. [1970]: 'On hidden-variable theories', *Journal of Mathematical Physics*, **11**, pp. 431–6.

HEYWOOD, P. AND REDHEAD, M. [1983]: 'Nonlocality and the Kochen-Specker paradox', *Foundations of Physics*, **13**, pp. 481–499.

HULTGREN, B. AND SHIMONY, A. [1976]: 'The lattice of verifiable propositions of the spin-1 system', *Journal of Mathematical Physics*, **18**, pp. 381–94.

KOCHEN, S. AND SPECKER, E. [1967]: 'The problem of hidden variables in quantum mechanics', *Journal of Mathematics and Mechanics*, **17**, pp. 59–88.

MACKEY, G. [1963]: *Mathematical Foundations of Quantum Mechanics*. Benjamin.

PEARLE, P. [1965]: 'A framework for hidden variable theories', preprint, Harvard University (unpublished).

SHIMONY, A. [1978]: 'Metaphysical problems in the foundations of quantum mechanics', ed. A. *International Philosophical Quarterly* **18**, pp. 3–17.

SHIMONY, A. [1981]: 'Critique of the papers of Fine and Suppes', *PSA 1980*, vol. 2, ed. A. ASQUITH AND R. GIERE (Philosophy of Science Association, E. Lansing, Michigan).pp. 572–80.

STEIN, H. [1972]: 'On the conceptual structure of quantum mechanics', *Paradigms and Paradoxes*, ed. R. Colodny (U. of Pittsburgh Press, Pittsburgh), pp. 367–438.

SWIFT, A. AND WRIGHT, R. [1980]: 'Generalized Stern-Gerlach experiments and the observability of arbitrary spin operators', *Journal of Mathematical Physics*, **21**, pp. 77–82.

VARADARAJAN, V. [1968]: *The Geometry of Quantum Theory*. Van Nostrand.

WIENER, N. AND SIEGEL, A. [1953]: 'A new form for the statistical postulate of quantum mechanics', *Physical Review*, **91**, pp. 1551–60.

Brit. J. Phil. Sci. 35 (1984), 47–53 *Printed in Great Britain*

Problems with Electoral Evaluations of Expert Opinions*

by ROY A. SORENSEN

An electoral evaluation of a set of expert opinions proceeds by treating the experts as voters. Although this method allows us to formalise our naive views about how to take expert advice, the formalisations are plagued by paradoxes which parallel those found in literature on social aggregation devices. This parallel suggests that our naive views about taking expert advice are in as much need of revision as our naive views about deriving group preferences from individual preferences.

Economists and political scientists have made progress by studying aggregation devices. An aggregation device is a function from a set of individual preferences or values into a set of collective preferences or values. For example, a group can decide whether Smith or Jones should hold office by using the simple majority rule. If more than half of the members vote for Smith, Smith is the group's choice. Although social aggregation devices have received the most attention, their utility has been shown in other areas. For example, in MacKay [1980] there is an interesting treatment of multi-event athletic competitions (like the decathlon) where the scoring system is analysed as an aggregation device which combines information about the competitors' performance in the individual events and yields an overall ranking. In this essay I shall be concerned with epistemic aggregation devices; functions from a set of individuals' epistemic preferences into a set of collective epistemic preferences. Some work has already been done on epistemic preferability,[1] but it has not been motivated by the problems I think epistemic aggregation devices can illuminate. We appear to evaluate expert opinions electorally. It is as if we assign votes to experts in accordance with their epistemic preferences, and then hold an election to determine what our epistemic preference should be. Knowledge by authority seems to be intimately connected to our use of epistemic aggregation devices. Just as social scientists have used paradoxes to show that our naive social aggregation devices require modification, I shall use analogues of these paradoxes to show that some of our naive epistemic aggregation devices require modification. The epistemic paradoxes included in this essay are transparent derivations from some prominent voting paradoxes. A reader who is familiar with voting theory will not find it difficult to derive other

---

* I am indebted to Joseph Hanna for a number of criticisms and corrections to a previous draft.

[1] The most notable work on epistemic preferability is found in Roderick Chisholm's epistemological writings. In 'A System of Epistemic Logic' (*Ratio*, vol. IV, no. 2, 1972, pp. 99–119) Chisholm and Robert G. Keim present 'The Calculus of Epistemic Preferability'.

Received 3 August 1982

epistemic paradoxes from some of the more sophisticated voting paradoxes. The ease of derivation indicates an underlying similarity between certain epistemic and voting principles. Although I concentrate on the problem of evaluating expert opinion, the reader should not conclude that this is the only kind of epistemic aggregation problem. For example, answering the question 'What is public opinion?' requires an appeal to epistemic aggregation devices given the assumption that public opinion depends on the opinions of members of the public. In this case, aggregation devices play a descriptive rather than a prescriptive role.

Much of our knowledge is based on authority. As we become increasingly specialised, we must often rely on experts. Although experts often disagree, we can nevertheless form justified opinions. In the face of conflicting expert opinions, we sometimes appeal to variations of the following principle:

(*M*) If the majority of experts prefer p over q, then p is epistemically preferable to q by an ideal advisee.

Of course, (*M*) has to be qualified. Since some experts are more reliable than others, a minority opinion is sometimes preferable. But even with a generous 'all other things equal' clause, (*M*) can lead to a breakdown of the transitivity of epistemic preferability. The possibility of cyclical majorities was first noted by Condorcet.

Table I

Voter	Preference scale for alternatives
1	(x_1, x_2, x_3)
2	(x_2, x_3, x_1)
3	(x_3, x_1, x_2)

Each voter has a transitive preference scale, but by majority rule, the social ordering is intransitive. A majority prefers x_1 to x_2, a majority prefers x_2 to x_3, and a majority prefers x_3 to x_1. The majorities are called cyclical because the choices can be ordered x_1, x_2, x_3, x_1, x_2, x_3, x_1. This voting paradox is turned into a challenge to the transitivity of epistemic preferability by letting the voters be experts and appealing to (*M*).

Cyclical majorities also show that the order in which issues are voted upon sometimes affects the outcome. If the voters are first asked to choose between x_1 and x_2, and second to choose between the result of that choice and x_3, x_3 is adopted. If the voters are first asked to choose between x_1 and x_3, and second to choose between the result of that choice and x_2, x_2 is adopted. By once again letting the voters be experts and assuming (*M*), it can be seen that the order of questioning can affect which answer is adopted. For example, suppose a patient asks his three doctors: 'Is diagnosis x_1 more likely to be correct than diagnosis x_2?' Two doctors answer affirmatively and one negatively. The patient therefore eliminates diagnosis x_2 and asks: 'Is diagnosis x_1 more likely to be correct than x_3?' Since two doctors favour x_3

over x_1, the patient fallaciously concludes that x_3 is the best supported diagnosis. If the patient had instead first asked 'Is diagnosis x_1 more likely to be correct than x_3?' and followed the same procedure of elimination, he would have concluded that x_2 is the best supported diagnosis. The patient's questioning procedure falsely presupposes that the losers of each pairwise comparison can be safely eliminated. The doctors' collective opinions can be inconsistent even when their individual opinions are consistent. Using (M), the collective opinion is 'x_1 is more likely than x_2, x_2 is more likely than x_3, and x_3 is more likely than x_1'. Since the collective opinion is inconsistent, the transitivity of the 'more likely to be correct' relation is not sufficient to insure safe elimination.

Cyclical majorities also arise in platform voting. Some voting procedures do not give the voters the opportunity to vote on single issues; they must instead choose between conjunctions of issues. Suppose the first issue is between A and $-A$, and the second is between B and $-B$.

Table II

Voter	Preference scale for alternatives			
1	$A \& B$	$A \& -B$	$-A \& B$	$-A \& -B$
2	$A \& -B$	$-A \& -B$	$A \& B$	$-A \& B$
3	$-A \& B$	$-A \& -B$	$A \& B$	$A \& -B$

As illustrated by the diagram below, appearing in Hillinger [1972], the majorities are cyclical and so the social preferences are intransitive.

Table III

$$\begin{tikzcd}
A \& B \arrow[r] \arrow[d] & A \& -B \arrow[d] \arrow[dl] \\
-A \& B \arrow[r] & -A \& -B \arrow[ul]
\end{tikzcd}$$

If separate votes were allowed, A would be preferred to $-A$ (by 1 and 2) and B would be preferred to $-B$ (by 1 and 3). Although majorities would prefer A and B under separate voting, $-A \& -B$ is preferred over $A \& B$ under platform voting (since the preferences of 2 and 3 constitute a majority). A platform only favoured by minorities can defeat a platform whose alternatives (considered separately) are favoured by a majority. Thus if the voters are treated as experts, and the issues as possible answers to the conjunctive question $(A \& B)$?, the voting anomaly is transformed into an epistemic anomaly. If one asks A? and B?, $(A \& B)$ is the best answer. But if one asks $(A \& B)$?, $(-A \& -B)$ is the best answer.

Of course, some voting procedure biases do not depend on cyclical majorities. One familiar phenomenon of roll-call voting is the bandwagon effect. Those that vote later than others sometimes vote in favour of a proposal because the prior votes indicate that it will pass and the voter hopes to be on the winning side. There is sometimes a snowballing effect because earlier bandwagon voters enhance the chance of passage by their votes so that later (and perhaps elsewise more reluctant) bandwagon voters are more apt to join. Thus roll-call voting can produce an artificial, lopsided victory. The victory is artificial in the sense that another voting procedure, say simultaneous voting, might produce a quite different result. Although experts sometimes bandwagon merely for the sake of being on the winning side, they can be motivated by more epistemically respectable reasons. As proponents of the Delphi technique and as stressed in Lehrer and Wagner [1981], experts should respect each other's opinions. An expert's epistemic preferences can be justifiably influenced by his knowledge of another expert's preferences. Yet this provides the basis for an epistemic bandwagon effect. For the sake of simplicity, suppose there are three highly respectful experts, 1, 2, and 3, who prior to the roll-call vote are respectively in favour, indifferent, and opposed to a proposition. However, they only learn the others' preferences by their votes. If the roll-call vote is taken in order 1, 2, 3, expert 1 votes in favour. Having learned that another expert favours the proposition, the opinion of 2 is swayed and he too votes in favour. Having learned that two experts favour the proposition, 3 reverses his opinion (since he is highly respectful) and the proposition is unanimously favoured. However, if the roll-call vote is taken in order 3, 2, 1, incremental disclosure of preferences and high respect results in the proposition being unanimously opposed. Although this example is oversimplistic, the general point holds for more complicated cases involving larger groups with different degrees of respect and irregular preference revelation. Disclosure order bias indicates that epistemic respect is trickier than has been supposed. An expert who knows that a large consensus is due to the preference snowballing effect should have less respect for it than one due to the independent preferences of the experts.

The analogy between elections and methods of evaluating expert opinion invites further political analogies. For example, voting theorists have appreciated the shrewdness of the Tamany Hall politician who said that he did not care who did the electing just as long as he did the nominating. Publishers and teachers can control discussion and research by controlling the dissemination of ideas. Controversies may rage but still within the limits they set in selecting which problems receive notoriety and which candidate solutions reach the community of inquirers. The direction of scientific growth can be affected in a more obvious way by the resource allocation decisions of scientific advisory committees. Here the relevance of work done on coalition formation, bargaining, and sophisticated voting is plain. Where political analogies are fruitful one should also expect game and decision theoretic notions to apply since they are fruitful in gaining an understanding of political notions. For instance, standardisation of terminology and notation can be viewed as a coordination problem. Since increased descriptive adequacy provides the basis for better prescriptions, there is the hope that attaining a better understanding of how communities of inquirers behave will result in good suggestions as to how their internal problems can be solved.

Recently another voting paradox was presented in Anscombe [1976]. A simpler presentation of the paradox appears in Gorman [1978]:

Table IV

	Questions		
Voter	*A*	*B*	*C*
1	1	1	1
2	1	1	1
3	1	−1	−1
4	−1	1	−1
5	−1	−1	1

Here the questions are settled by yes–no majority vote but a majority of the voters vote in the minority on a majority of the questions. Although each of the questions is answered affirmatively by a 3 to 2 majority, voters 3, 4, and 5 are unsatisfied with the overall outcome in the sense that each votes unsuccessfully on a majority of the questions. Thus in one sense majority voting has led to the frustration of the will of the majority. As shown in Sorensen [1982], the size of this unsatisfied majority can be arbitrarily close to 100 per cent of those voting. Gorman argues that the crucial issue raised by Anscombe is that

> . . . we have two possible and conflicting criteria for saying 'the majority has its will fulfilled' . . . should there be more successful votes than unsuccessful votes for each question? Or should there be more successful votes than unsuccessful votes for the majority of voters.[1]

We cannot always maximise satisfaction of voters and satisfaction of their desires simultaneously.

This maximisation dilemma undermines the following principle of collection:

(*C*) If a majority of experts prefer a majority of the conjuncts of a conjunction *C* over their negations, then *C* is epistemically preferable to the conjunction of the negations of the conjuncts of *C* by an ideal advisee.

In symbols, where '${}_{p_1}p_{-p_1}$' reads 'p_1 is epistemically preferred to $-p_1$ by a majority of the experts' and '${}_{p_1}p_{a-p_1}$' reads (p_1 is epistemically preferred to $-p_1$ by an ideal advisee':

$$({}_{p_1}p_{-p_1} \;\&\; {}_{p_2}p_{-p_2} \;\&\; \ldots \;\&\; {}_{p_n}p_{-p_n}) \supset$$
$$((p_1 \;\&\; p_2 \;\&\; \ldots \;\&\; p_n)p_a(-p_1 \;\&\; -p_2 \;\&\; \ldots \;\&\; -p_n))$$

---

[1] J. L. Gorman, 'A Problem in the Justification of Democracy', *Analysis*, 38 [1978], p. 48.

If the conjuncts are distributively preferred by the experts, then the conjunction is preferred by an ideal advisee. Principle (*C*) should be qualified in a number of ways. If the favourable preferences of each expert are of weak intensity and the unfavourable preferences are of a sufficiently strong intensity, the minority conjunction is preferred. If the conjuncts of *C* are jointly inconsistent, *C* could again fail to gain preference. But it does seem that a generous *ceteris paribus* clause could smooth over these rough spots. However, Gorman's table provides a basis for the construction of a strong counterexample to (*C*). Since *p* is a degenerate conjunction (being equivalent to *p* & *p*), (*C*) implies a weaker version of (*M*):

(*M′*) If a majority of experts prefers *p* over −*p*, then *p* is epistemically preferable to −*p* by an ideal advisee.

Let the voters in Gorman's table be equally reliable experts expressing equally intense preferences with respect to the possible answers to three independent questions. Given that the ideal advisee has no outside evidence supporting any answer over another, it seems appropriate that he rely on the experts. Since $AP-A$ & $BP-B$ & $CP-C$, the advisee infers AP_a-A & BP_a-B & CP_a-C by (*M′*). Since a majority prefers a majority of negative answers, he infers $(-A \,\&\, -B \,\&\, -C)\ P_a\ (A \,\&\, B \,\&\, C)$ by (*C*). Thus if the advisee accepts (*C*) he must reject even a charitable interpretation of

$$({}_{p_1}p_{a-p_1} \,\&\, {}_{p_2}p_{a-p_2} \,\&\, \dots \,\&\, {}_{p_n}p_{a-p_n}) \supset$$
$$-((-p_1 \,\&\, -p_2 \,\&\, \dots \,\&\, -p_n)p_a(p_1 \,\&\, p_2 \,\&\, \dots \,\&\, p_n))$$

The maximisation dilemma shows how it is possible for someone to abide by the majority opinions of his advisers on each and every question and yet have a majority disagree with a majority of his answers. By using (*C*) an advisee can be led to inconsistency even though his advisers are distributively consistent. Majoritarean decision procedures are vulnerable to the maximisation dilemma because they do not reflect the preference intensities of the collective. By ignoring marginal differences one makes grouping arrangements relevant. For example, suppose a teacher grades his students on the basis of three pass/fail, true/false tests. If a student correctly answers a majority of a test's questions, he passes the course. Students Art and Bob perform as follows:

Table V

	Tests				Tests		
	A	*B*	*C*		*A*	*B*	*C*
1	1	1	−1	1	−1	−1	1
2	1	1	−1	2	−1	−1	1
3	−1	−1	−1	3	1	1	1

Art and Bob answer oppositely. Overall Bob answers a majority of the questions correctly but fails a majority of the tests. Overall Art answers a majority of the questions incorrectly but passes a majority of the tests. Therefore, Art passes and Bob fails the course even though Bob answered more questions correctly than Art. This outcome is due to the fact that when Art passed a test he passed by a narrow margin and when he failed he failed by a large margin, while Bob's margins had the opposite character. This failure to reflect marginal differences provides the basis for gerrymandering and for criticisms of multi-stage, winner-take-all aggregations of regional electoral results like the American electoral college.

Although voting theorists tend to concentrate on problems of collective rationality, the problems infect individual rationality as well. As emphasized in this essay, infection occurs by means of the epistemic authority collectives carry (which is not to deny that infection occurs by other kinds of authority collectives have). There are other significant means of infection. For example, we often make judgements as to the truth value of a proposition or the applicability of an expression partly or wholly on the basis of criteria which are neither necessary nor sufficient conditions. The criteria behave like (sometimes weighted) voters since the decision is majoritarean. Like any other aggregation problems, this sort of criterial judgement is vulnerable to subtle failures of consistency.

Michigan State University

REFERENCES

ANSCOMBE, G. E. M. [1976]: 'On Frustration of the Majority by the Fulfillment of the Majority's Will', *Analysis*, **36**, pp. 161–8.

GORMAN, J. L. [1978]: 'A Problem in the Justification of Democracy', *Analysis*, **38**, pp. 46–50.

HILLINGER, CLAUDE [1972]: 'Voting on Issues and on Platforms', *Behavioral Science*, **16**, pp. 564–6.

LEHRER, KEITH and WAGNER, CARL [1981]: *Rational Consensus in Science and Society*. D. Reidel.

MACKAY, ALFRED F. [1980]: *Arrow's Theorem: The Paradox of Social Choice*. Yale University Press.

SORENSEN, ROY A. [1982]: 'Democratic Frustration of Vast Majorities', *Political Studies*, **30**, pp. 272–3.

Brit. J. Phil. Sci. 35 (1984), 55–64 *Printed in Great Britain*

Discussions

PROBABILISTIC CAUSALITY AND PREEMPTION

Suppes has proposed a probabilistic theory of causality.[1] This theory has certain advantages over rival non-probabilistic accounts, *e.g.*, coping with cases in which the causes are less than sufficient for their effects. In order to be considered fully adequate, however, a probabilistic theory must also be able to deal with various other problem cases. In this paper, I will argue that Suppes' theory cannot successfully account for one such type of case involving preemption.

Suppes' Theory

Suppes' definition of a 'cause' presupposes two other notions: '*prima facie* cause' and 'spurious cause'.

A_t is a *prima facie* cause of $B_{t'}$ if and only if

(i) $t < t'$
(ii) $P(A_t) > 0$
(iii) $P(B_{t'} \mid A_t) > P(B_{t'})$[2]

Clause *(i)* requires that every *prima facie* cause precede its *prima facie* effects, and, thus, establishes the asymmetry of the causal relation. Clause *(ii)* states that there must be some probability that the cause event occurs. And most importantly, according to clause *(iii)*, the probability of $B_{t'}$ conditional on A_t must be greater than the probability of $B_{t'}$ simpliciter.

Not every *prima facie* cause, however, is a genuine cause. For example, a specific barometer reading may increase the probability of a storm, but the barometer reading obviously does not cause the storm. The former is a 'spurious cause' of the latter. Suppes gives various senses to 'spurious cause', but the following definition will serve for the purposes of this paper:

> A is a spurious cause of B if and only if A is a *prima facie* cause of B and there is an event C, earlier than A, such that
>
> *(i)* $P(A, C) > 0$
> *(ii)* $P(B \mid A, C) = P(B \mid C)$
> *(iii)* $P(B \mid A, C) \geqslant P(B \mid A)$[3]

---

[1] Patrick Suppes [1970]: *A Probabilistic Theory of Causality*, Amsterdam: North Holland Publishing Co.

[2] *Ibid.*, p. 12.

[3] *Ibid.*, p. 23.

Clause (*i*) establishes that the probability of A and C occurring together is greater than zero. Clause (*ii*) states that the probability of B given A and C is equal to the probability of B given C. In our example, the probability of the storm given the barometer reading and earlier weather conditions is equal to the probability of the storm given those earlier weather conditions. Finally, clause (*iii*) has it that the probability of B given both A and C is equal to or greater than the probability of B given A. Clearly, the probability of the storm given both the barometer reading and the weather conditions is equal to or greater than the probability of the storm given the barometer reading. The barometer may, for instance, malfunction indicating a storm when weather conditions are not ripe for one. A *genuine cause*, according to Suppes, is a *prima facie* cause which is not spurious.

Preemption

Suppose that Allen and Cathy decide to try to kill Bob. Cathy and Allen separately connect electrical wires to Bob's chair, which he sits in each day at the same time. As it happens, Cathy and Allen use the same switching mechanism on the outside of Bob's house for their electrical connections. This mechanism operates in such a way that only current from one source can flow through at a given time. Now suppose the probability of Bob's death, given that Allen pulls his switch in time for the current from his source to pass through the switching mechanism, is 1/2. In turn, if Cathy pulls her switch the probability of Bob's death is 1/2. Cathy's switch is much farther away from the switching mechanism than is Allen's switch so that the current from her source takes longer to reach the chair. The following scenario unfolds: Cathy's ($C_{t''}$) pulls her switch slightly earlier than does Allen ($A_{t'}$). The current from Allen's source passes through the switching mechanism and breaks the circuit before Cathy's current reaches the mechanism. Bob dies of electrical shock (B_t) from Allen's current. Thus $A_{t'}$ causes B_t, and the causal chain running from $A_{t'}$ preempts $C_{t''}$ in causing B_t.

Given Suppes' definition of 'spurious cause', however, we must say that $C_{t''}$ screens off $A_{t'}$ from B_t, and thus that $A_{t'}$ is a spurious cause of B_t. This can be seen as follows: First, $A_{t'}$ is a *prima facie* cause of B_t since the probability of Bob's death given that Allen pulls his switch is greater than the probability of Bob's death simpliciter. The probability of both $A_{t'}$ and $C_{t''}$ occurring is greater than zero, and thus clause (*i*) of the definition is satisfied. Clause (*ii*) is also fulfilled since the probability of Bob's death given that both Allen and Cathy pull their switches is equal to the probability of Bob's death given that Cathy pulls her switch. This is so since the probability of B_t given $C_{t''}$ is 1/2 as is the probability of B_t given both $A_{t'}$ and $C_{t''}$. The probability of B_t given $A_{t'}$ and $C_{t''}$ is 1/2 since $A_{t'}$ insures that the current from Cathy's source will not flow through the switching mechanism, and thus the probability of B_t given both $A_{t'}$ and $C_{t''}$ is simply equal to the probability of B_t when $A_{t'}$ occurs without $C_{t''}$ (*i.e.* 1/2). In turn the probability of Bob's death given that Cathy pulls her switch is 1/2. This is so since if Cathy pulls her switch and Allen does not, the probability of death is 1/2, and if Cathy pulls her switch and Allen does, then the probability of death is 1/2. Hence $P(B_{t'}|A_{t'}, C_{t''}) = P(B_t|C_{t''})$. Clause (*iii*) is also met. We have shown that the probability of B_t given $A_{t'}$ and $C_{t''}$ is 1/2, and thus equal to the probability of B_t and $A_{t'}$, which is also 1/2. Hence $P(B_{t'}|A_{t'}, C_{t''}) \geqslant (B_t \,|\, A_{t'})$. From this it can be seen by the definition of 'spurious cause' that $A_{t'}$ is *not* a genuine cause of B_t since there exists an event $C_{t''}$, earlier than $A_{t'}$, which 'screens off' $A_{t'}$ from B_t. This result, however, runs contrary to our assumption that $A_{t'}$ preempts $C_{t''}$ in causing B_t.

DOUGLAS EHRING
Southern Methodist University, Dallas

ARE PARADIGMS INCOMMENSURABLE?*

Thomas Kuhn [1970] and others, including Paul Feyerabend [1975] and Barry Barnes [1982], have argued that there can be no comparison of competing paradigms based solely on experimental evidence. 'There is no appropriate scale available with which to weigh the merits of alternative paradigms: they are incommensurable' [Barnes [1982], p. 65]. Briefly stated, the argument is as follows. There can be no neutral observation language. All observation terms are theory laden and thus we cannot compare experimental results because in different paradigms terms describing experimental results have different meanings, even when the words used are the same.[1] An example would be the term 'mass' which in Newtonian mechanics is a constant, while in Einstein's special theory of relativity it depends on velocity. I believe this view is incorrect.

---

* This material is based upon work supported by the National Science Foundation under Grant No. SES-8204074. Any opinions, findings, conclusions or recommendations expressed in this publication are those of the author and do not necessarily reflect the views of the National Science Foundation.

[1] Kuhn offers the following statement on incommensurability. 'These examples point to the third and most fundamental aspect of the incommensurability of competing paradigms. In a sense that I am unable to explicate further, the proponents of competing paradigms practice their trades in different worlds. . . . Equally it is why, before they can hope to communicate fully, one group or the other must experience the conversion that we have been calling a paradigm shift. Just because it is a transition between incommensurables, the transition between competing paradigms cannot be made a step at a time, forced by logic and neutral experience' ([1970], p. 150). 'This need to change the meaning of established and familiar concepts is central to the revolutionary impact of Einstein's theory' ([1970], p. 102). Kuhn has recently modified his view. In an address to the Philosophy of Science Association (Oct. 1982) he spoke of 'local incommensurability', which I take to mean the difficulty of context dependent translation Feyerabend is even harder to pin down. At times he states that theories are incommensurable only if they are not interpreted in an 'independent observation language' ([1975], p. 274). Elsewhere he seems to deny the possibility of such a language. 'Adopting the point of view of relativity we find that the experiments, *which of course will now be described in relativistic terms*, (italics in original) are *relevant* to the theory, and we also find that they *support* the theory. Adopting classical mechanics we again find that the experiments *which are now described in the very different terms of classical physics* are relevant, but we also find that they undermine classical mechanics' ([1975], p. 282).

In this note I will demonstrate that an experiment, described in procedural, theory-neutral (between the two competing paradigms or theories) terms, gives different theory-neutral results when interpreted in the two alternative paradigms. Thus a measurement of the quantity derived will unambiguously distinguish between the two. I will take as an example one of Kuhn's own exemplars, the difference between Newtonian and Einsteinian mechanics. The case in point can be loosely described as the scattering of equal 'mass' objects.

The experimental procedure is as follows. Consider a class of objects, let us say billiard balls.[1] The objects are examined pairwise by placing a compressed spring between them. The spring is allowed to expand freely and the velocities of the two objects measured. Because we restrict ourselves to a single frame of reference in the laboratory the measurement is theory-neutral. We then select two balls whose velocities are equal. Of course, a Newtonian would interpret this as giving two objects for which the Newtonian masses M_N are both equal and constant, while an Einsteinian would say that the relativistic masses $M_R = M_{0R}\sqrt{1 - V^2/C^2}$ are equal and thus that the rest masses M_{0R} are equal. This is agreed, but the point is that the procedure itself is theory-neutral. One of the objects is placed at rest in the laboratory and the other given a velocity V_1 (again theory-neutral) and the particles allowed to scatter off each other. Care is taken to make the collision elastic, *i.e.* no energy of any sort is emitted. The final velocities of the objects V_2 and V_3, respectively, are measured as is the angle between these velocities. It is again true that the assignment of momenta and energies to the particles will be different but the measurement of the angle does not depend on these assignments. As is well-known, and demonstrated below, the predicted value of this angle differs in Newtonian and relativistic mechanics. For a Newtonian $\theta_N = 90°$, while for an Einsteinian $\theta_R < 90°$. In the calculation below we shall use units such that $C = 1$ (the velocity of light). We shall also assume for calculational convenience that the two outgoing velocities are equal. This apparent loss of generality is unimportant, although the proof can be generalised.[2] The fact that *any single* experiment can distinguish between the two paradigms is enough to show that they are commensurable.

---

[1] The analysis also seems to apply to proton-proton or electron-electron scattering. An anti-realist might object, however, that protons and electrons are theoretical entities whose definition includes mass. Thus the two views of the experiments would be incommensurable because they are talking about different entities. The advantage of billiard balls is that they can be defined ostensively. You can hold one in your hand or point to it.

[2] Dr Michael Redhead has graciously provided such a general proof. Let θ^* and ϕ^* be the scattering angles in the centre of mass system (Note: $\theta^* + \phi^* = 180°$) and let θ, ϕ be the angles in the laboratory system. Then $\tan\theta = \sqrt{2/\gamma + 1}\,\tan\frac{1}{2}\theta^*$ where $\gamma = 1/\sqrt{1 - V_1^2/c^2}$, and similarly for ϕ. Then it can be shown quite easily that $\tan(\theta + \phi) = (2\sqrt{\gamma + 1}/\gamma - 1)/\sin\theta^*$ so that except for $\theta^* = 0°$ $\tan(\theta + \phi) < \infty$ or $\theta + \phi < 90°$.

Let us first consider the experiment from a Newtonian point of view. Let $\vec{P}_{1N}$, $\vec{P}_{2N}$, $\vec{P}_{3N}$ be the initial and final momenta and E_{1N}, E_{2N}, E_{3N} be the initial and final energies. We further note that in this case $E_{IN} = P^2_{IN}/2M_N$. We now apply the Newtonian of conservation of momentum.

$$\vec{P}_{1N} = \vec{P}_{2N}+\vec{P}_{3N},$$

$$P^2_{1N} = P^2_{2N}+P^2_{3N}+2P_{2N}P_{3N}\cos\theta_N.$$

This gives $\cos\theta_N = (P^2_{1N}-P^2_{2N}-P^2_{3N})/2P_{2N}P_{3N}$ (1)

Applying Newtonian conservation of energy we have

$E_{1N} = E_{2N}+E_{3N}$ or

$$P^2_{1N}/2M_N = P^2_{2N}/2M_N+P^2_{3N}/2M_N.$$

This gives $P^2_{1N}-P^2_{2N}-P^2_{3N} = 0$ (2).

Combining (1) and (2) we get the result

$$\cos\theta_N = 0 \quad \text{or} \quad \theta_N = 90^\circ, \quad \text{because } P_{2N} \text{ and } P_{3N} \neq 0.$$

Let us now consider the experiment interpreted in relativistic terms. Using relativistic conservation of momentum we have $\vec{P}_{1R} = \vec{P}_{2R}+\vec{P}_{3R}$ and we obtain by a calculation identical to that above

$$\cos\theta_R = (P^2_{1R}-P^2_{2R}-P^2_{3R})/2P_{2R}P_{3R}.$$

We may rewrite this using relativistic energy and momentum as

$$\cos\theta_R = (E^2_{1R}-M^2_{0R}-E^2_{2R}+M^2_{0R}-E^2_{3R}+M^2_{0R})/2P_{2R}P_{3R}$$

$$\cos\theta_R = (E^2_{1R}-E^2_{2R}-E^2_{3R}+M^2_{0R})/2P_{2R}P_{3R} \tag{3}$$

Using relativistic conservation of energy we have

$$E_{1R}+M_{0R} = E_{2R}+E_{3R}. \tag{4}$$

Squaring both sides gives

$$E^2_{1R}+2E_1M_{0R}+M^2_{0R} = E^2_{2R}+2E_{2R}E_{3R}+E^2_{3R} \tag{5}$$

Rewriting we have

$$E^2_{1R}-E^2_{2R}-E^2_{3R}+M^2_{0R} = 2E_{2R}E_{3R}-2E_1M_{0R} \tag{6}$$

We can now show that the expressions given in Eq. (6), which are the numerator in the expression for $\cos\theta_R$ (Eq. (3)) are greater than zero and this $\cos\theta_R > 0$ or $\theta_R < 90^\circ$.

Consider $2E_{2R}E_{3R}-2E_{1R}M_{0R}$. We now consider the case that $E_{2R} = E_{3R} = (E_{1R}+M_{0R})/2$ from Eq. (4).

We now have

$$
\begin{aligned}
&2((E_{1R}+M_{0R})/2)^2-2E_{1R}M_{0R}\\
&=2(E_{1R}^2+2E_{1R}M_{0R}+M_{0R}^2)/4-2E_{1R}M_{0R}\\
&=E_{1R}^2/2+E_{1R}M_{0R}+M_{0R}^2/2-2E_{1R}M_{0R}\\
&=E_{1R}^2/2-E_{1R}M_{0R}+M_{0R}^2/2=((E_{1R}-M_{0R})/\sqrt{2})^2
\end{aligned}
$$

By the condition of our experiment $E_{1R} > M$ and thus this expression is greater than 0.

Thus $2E_{2R}E_{3R}-2E_{1R}M_{0R} = ((E_{1R}-M_{0R})/\sqrt{2})^2$ in this case but $2E_{2R}E_{3R} -2E_{1R}M_{0R} = E_{1R}^2-E_{2R}^2-E_{3R}^2+M_{0R}^2$ and must be greater than 0 and hence $\cos\theta_R > 0$ and $\theta_R < 90°$, because $P_{2N} = P_{3R} \neq 0$.

We have thus shown that in this procedurally defined, theory-neutral, experiment the predicted results when interpreted within the two paradigms give unambiguous and different results *i.e.* $\theta_N = 90°$, $\theta_R < 90°$. Thus a measurement of the angle between the velocities of the two outgoing objects will clearly distinguish between the two paradigms and the paradigms are commensurable. The fact that two paradigms are commensurable does not require a scientist to accept the one consistent with the experimental results, namely relativity. A clever classical physicist might very well accommodate the results by postulating new mechanical effects depending on motion through the ether.[1] These two issues of commensurability and theory choice are often conflated, as exemplified in the statement by Barnes cited earlier, but it is useful to keep them separate and distinct.

ALLAN FRANKLIN
University of Colorado

REFERENCES

Barnes, B. [1982]: *T. S. Kuhn and Social Science*. Macmillan.
Feyerabend, P. [1975]: *Against Method*. New Left Books.
Kuhn, T. S. [1970]: *The Structure of Scientific Revolutions*. Chicago University Press, 2nd ed.

EINSTEIN AND SPECIAL RELATIVITY: WHO WROTE THE ADDED FOOTNOTES?

A recent note in this *Journal* by Cullwick [1981] attributes a certain footnote in the English translation of Einstein's original paper on special relativity (published in Lorentz *et al.* [1923]) to Sommerfeld. A recent book by Miller [1981] attributes to Sommerfeld all those footnotes that appear in the English translation but not in the original paper.

---

[1] I am grateful to the anonymous referee for making this point.

Unfortunately Cullwick and Miller have not presented any evidence for attributing those footnotes to Sommerfeld, and the only evidence for their position appears to be the inscription 'With notes by A. Sommerfeld' on the title page of the collection (Lorentz *et al.* [1923]). The inscription obviously refers to a section of the book headed 'Notes by A. Sommerfeld', comprising notes on Minkowski's paper which is included in the collection; there is no specific evidence that the inscription on the title page refers also to the footnotes added to Einstein's paper. The first German edition of the collection (Lorentz *et al.* [1913]) shows a similar situation: there is a corresponding inscription on the title page, but no specific evidence to relate the added footnotes to Sommerfeld. The fourth German edition (Lorentz *et al.* [1922]), a somewhat different collection, from which the English translation was made, provides no further evidence on that point.

When a revision of a paper is published in the author's lifetime, it seems reasonable to assume that any alterations or additions are the author's own, unless there is explicit information to the contrary. This seems to apply especially to the first footnote to the English translation of Einstein's paper, which reads: 'The preceding memoir by Lorentz was not at this time known to the author.' Since that information must have been provided by Einstein himself, it seems implausible to attribute the authorship of that footnote to Sommerfeld.

In the case of Minkowski the situation was different: he was the only one of the authors concerned who was no longer living when the German collection was being published in 1913; that fact would explain the appropriateness of Sommerfeld's writing notes on his paper rather than the others.

Otto Blumenthal's Foreword to the 1913 edition (not reproduced in the English translation) acknowledges the friendly co-operation (*freundliche Bereitwilligkeit*) of Lorentz and Einstein, which made it possible to publish that collection. (They were the only authors, apart from Minkowski, to have papers included in that edition.) It seems reasonable to assume, unless there is clear evidence to the contrary, that Einstein's co-operation included the addition of the footnotes in question.

IAN MCCAUSLAND
University of Toronto

REFERENCES

CULLWICK, E. G. [1981]: 'Einstein and Special Relativity: some inconsistencies in his electrodynamics'. *British Journal for the Philosophy of Science*, **32**, pp. 167–76.

LORENTZ, H. A. *et al.* [1923]: *The Principle of Relativity*. Methuen.

LORENTZ, H. A. *et al.* [1913, 1922]: *Das Relativitätsprinzip*. Teubner.

MILLER, A. I. [1981]: *Albert Einstein's Special Theory of Relativity: Emergence (1905) and Early Interpretation (1905–1911)*. Massachusetts: Addison-Wesley.

A GEDANKENEXPERIMENT TO MEASURE THE ONE WAY VELOCITY OF LIGHT

As is well known, many authors have maintained that all the predictions concerning observable phenomena within the Special Theory of Relativity (STR) can be obtained by assuming that the one way velocity of light (OWVL) is not c, provided only the round trip velocity of light is still maintained to be c. This thesis, known as the Conventionality of Simultaneity (CS), thus entails that one cannot measure OWVL. (In his 1905 STR paper Einstein merely *assumed* OWVL $= c$.) Fung and Hsieh [1980] and Nissim-Sabat [1982] have shown one can measure OWVL if OWVL $\neq c/(1-(1-2\varepsilon)\cos\theta)$, $0 \leq \varepsilon \leq 1$, with θ being measured with respect to a direction in space. (Fung and Hsieh had also proposed a method for measuring OWVL when OWVL $= c/(1-(1-2\varepsilon)\cos\theta)$ but Nissim-Sabat has shown this method not to be correct.) Furthermore, Winnie [1970] and others have shown that an asymmetry in OWVL entails an asymmetry in all one way velocities.

Most recently, Kolen and Torr [1982] have proposed a method for measuring OWVL, one which would be within the scope of the CS thesis and thus of STR. They propose to have two clocks A and B set a distance d apart, and to compare measurements of light–travel times taken at a given time with measurements taken 12 hours later, the rotation of the earth interchanging the relative positions of the two clocks. Without going into further detail, one can see that the interpretation of such measurements within STR and especially within the CS thesis is fraught with difficulties. The concept of a rigid body is one which cannot be easily incorporated in STR, especially when one has the rigid body rotating. Within the CS thesis, this concept is even more suspect, as shown by Townsend [1980] in his critique of the work of Jackson and Pargetter [1977]. Townsend shows that one cannot assume that the body is merely "rigid enough" without assuming that CS is false. Furthermore, even if one were to accept Kolen and Torr's tacit assumptions about the earth's rigidity and one were to neglect the earth's revolution about the sun, one is still faced with another difficulty. If one performs the experiment by measuring 12 hours using an inertial clock, say one on the sun, one has no guarantee that the phase difference between two clocks has remained the same. Their histories are slightly different: if clock A is a distance d East of B then, as the earth turns, A travels a distance d farther in the westward direction than does B while B travels a distance d farther in the eastward direction. Given the possible anisotropy of space inherent in the CS thesis, Kolen and Torr have to show that the two clocks will experience an equal lapse of time.

Yet one can devise a Gedankenexperiment which does *not* utilise the rotation of the earth in order to interchange the positions of the two clocks but one which would maintain some features of the experiment proposed by Kolen and Torr. Then one can demonstrate that this experiment would indeed succeed in measuring OWVL. Let us have two identical isochronous clocks A and B located at the points α and β on the x-axis, with $d = x_\beta - x_\alpha$. Following Kolen and Torr we have an interval timer located at α. Let it be agreed that the two clocks will each send a signal when each reads a specific time, say 10 am. The interval timer will be started when the pulse from A arrives and will be stopped when the pulse from B arrives. Following Reichenbach, let OWVL $= c/2\varepsilon$ $(0 \leq \varepsilon \leq 1)$ from α to β (along the positive x direction) and OWVL $= c/[2(1-\varepsilon)]$ from β to α (along the negative x direction). Then the interval timer will measure the time difference between the two pulses to be

$$\tau_1 = \Delta T + 2d(1-\varepsilon)/c,$$

ΔT being the time difference between the two clocks attributable to the fact that they cannot be synchronised; that they do not necessarily read 10 am 'at the same time'.

Let us have each of the clocks undergo the following as soon as it has sent out the 10 am pulse:

(1) accelerate to a *round trip velocity* (RTV), v, in a time t as measured by the clocks. (This is allowed by the CS thesis.)
(2) go to the point γ on the x-axis with that same RTV v $(x_\gamma > x_\beta > x_\alpha)$
(3) at γ come to a full stop with a deceleration time t
(4) turn around and accelerate in a time t to a RTV v in the negative x direction
(5) clock A stops at β and clock B at α with the same deceleration time t
(6) each clock sends a light pulse to the interval timer as soon as it has stopped

For the sake of simplicity we can consider the limit $t \to 0$. Note that the two clocks have had exactly the same history (*e.g.*, they suffer exactly the same accelerations) except that A went from α to β with a one way velocity $\overrightarrow{v}_\varepsilon$, while B went from β to α with a one way velocity $\overleftarrow{v}_\varepsilon$. According to Winnie,

$$\overrightarrow{v}_\varepsilon = cv/(c+v(2\varepsilon-1))$$
$$\overleftarrow{v}_\varepsilon = cv/(c-v(2\varepsilon-1)).$$

Winnie shows clock A would measure the time from α to β to be (p. 97):

$$\tau_{\alpha\beta} = d(1-v^2/c^2)^{\frac{1}{2}}/v. \tag{1a}$$

The same as under standard synchronisation! Clock B would measure the time from β to α to be again

$$\tau_{\beta\alpha} = d(1-v^2/c^2)^{\frac{1}{2}}/v = \tau_{\alpha\beta}. \tag{1b}$$

The two clocks would measure exactly the same time τ_γ for the segment $\beta \to \gamma \to \beta$ of their trip. Thus each clock would send out a pulse at a time $\tau_{\alpha\beta} + \tau_\gamma$ later than it sent out the original ('10 am') pulse. The interval timer would receive the pulse from B immediately but that from A a time

$2d(1-\varepsilon)/c$ after emission so that the time difference between the pulse from A and that from B would be

$$\tau_2 = \Delta T - 2d(1-\varepsilon)/c.$$

Thus one can measure ε:

$$\varepsilon = 1 - c(\tau_1 - \tau_2)/4d$$

(Inasmuch as the proposed experiment hinges on the validity of equations ($1a$) and ($1b$), it must be pointed out that we have ample experimental confirmation of their validity. These equations predict the same decay time (and, therefore, the same decay length in the laboratory) for unstable particles, say pions, of a given RTV regardless of the direction of motion: this is indeed what has been observed in countless experiments, although, to this author's knowledge, no systematic study to this effect has ever been performed.)

CHARLES NISSIM-SABAT
Northeastern Illinois University

REFERENCES

FUNG, S.-F. and HSIEH, K. C. [1980]: 'Is the Isotropy of the Speed of Light a Convention?', *American Journal of Physics*, **48**, pp. 654–7.

JACKSON, F. and PARGETTER, R. [1977]: 'Relative Simultaneity in the Special Theory of Relativity', *Philosophy of Science*, **44**, pp. 464–74.

KOLEN, P. and TORR, D. G. [1982]: 'An Experiment to Measure the One Way Velocity of Propagation of Electromagnetic Radiation', *Foundations of Physics*, **12**, pp. 401–11.

NISSIM-SABAT, C. [1982]: 'Can One Measure the One Way Velocity of Light?', *American Journal of Physics*, **50**, pp. 533–6.

TOWNSEND, B. [1980]: 'Jackson and Pargetter on Distant Simultaneity', *Philosophy of Science*, **47**, pp. 646–55.

WINNIE, J. A. [1970]: 'Special Relativity Without One Way Velocity Assumptions, Part I', *Philosophy of Science*, **37**, pp. 81–99.

Review Article

AN UNREAL IMAGE*

1 *Between Realism and Empiricism?*
2 *Syntax or Semantics?*
3 *What Does it Take to 'Save the Phenomena'?*
4 *Is Causal Explanation 'Context Dependent'?*

Bas van Fraassen's *The Scientific Image* has rightly been the focus of a good deal of recent philosophical attention. It contains challenging theses and clear arguments. One of its virtues, therefore, is that it provides a great deal to disagree with.

The central purpose of the book is to build and defend a half-way house between realism and empiricism. I shall first describe the main features of this compromise and then explain why I see no reason to buy into it (section 1). Van Fraassen's whole approach to scientific theories is premised on the 'semantic' characterisation of them as sets of models rather than sets of sentences. He insists that this characterisation opens up many insights which are not to be had from the usual 'syntactic' viewpoint. In section 2 I shall explain why I remain unconvinced of the advantages of the semantic approach. In section 3 I take up one point of detail—van Fraassen's inadequate characterisation of the notion of 'empirical adequacy'. In the course of pursuing his central purpose van Fraassen advances several interesting theses concerning more particular issues. In section 4 I take up one of these: his thesis that the notions of cause and of causal explanation are importantly and essentially 'context-dependent'. My comments will be restricted to these four topics even though this means leaving out of account some interesting features of the book—notably the treatment of probability in chapter 6.

1 BETWEEN REALISM AND EMPIRICISM?

According to van Fraassen's characterisation, *scientific realism* asserts (*i*) that the languages in which scientific theories are expressed are to be interpreted literally—talk of elementary particles, curved spacetime and the rest is not to be watered-down to mere 'as if' talk; (*ii*) that scientific theories are true-or-false attempted descriptions of the world—and true-or-false again not in some watered-down sense but in the fully-fledged correspondence sense; and (*iii*) that acceptance of a scientific theory involves the belief that the theory is true (although this belief may be tentative and hedged). *Scientific empiricism* on the other hand, asserts that in so far as theories transcend all possible observational data they are not to be interpreted literally—theoretical entities irreducible to observational data are to be interpreted as fictions, codification schemes or whatever. Indeed some empiricists, of course, still insist that no real sense can be made of observation-transcendent truth.

---

* Review of BAS C. VAN FRAASSEN [1980]: *The Scientific Image*. Clarendon Press. Pp. xi+235.
I completed this review while holding a research fellowship at the Center for Philosophy of Science at the University of Pittsburgh. I should like to thank all those who made my stay there such an enjoyable and stimulating one. I received valuable comments on an earlier draft from my fellow visiting fellows Ron Giere and Ron Laymon and also from my colleagues Colin Howson and Peter Urbach.

Van Fraassen rejects both realism and this sort of empiricism and constructs instead an unusual middle position, which he calls 'constructive empiricism'.[1] This says first that the languages in which scientific theories are expressed should be interpreted literally:

> If the theory's statements include 'There are electrons', then the theory says that there are electrons (p. 11).

In other words, no positivistic reduction of science is countenanced: semantic scruples about whether sense can really be made of statements which transcend in principle all possible data are to be disregarded; theories are indeed true-or-false (attempted) descriptions of reality.[2] *On the other hand*, there is, according to constructive empiricism, no reason to assume that what even our best accepted theories tell us is true—nor even that the *aim* of science is to produce true theories. Acceptance of a theory involves, *not* the belief that it is true, but only the belief that it is 'empirically adequate'[3]—that it 'saves the phenomena'. It is important to note that this is taken to involve *all* the phenomena—past, present, and future, observed and unobserved. Hence we can never know for sure that a theory is empirically adequate. Nevertheless, acceptance is taken to involve an, at any rate tentative, belief in empirical adequacy *and* to involve further a 'certain commitment to confront any future phenomena by means of the conceptual resources of this theory' (p. 12). It has often been argued that if empiricism were widely adopted by scientists then science would stagnate: the heuristic impetus supplied by a realist outlook would be lacking. This 'commitment' concerning future phenomena is intended to supply the potentially missing heuristic force.

What arguments might incline us to adopt van Fraassen's compromise 'constructive empiricism'? His main tactic is to argue that the view makes sense of science and is, moreover, the *minimal* view which does so; therefore to adopt any stronger, realist position would involve assumptions which are otiose. So he spends a good deal of time countering various arguments to the effect that some undoubted feature of science *entails* a realist view of theories. Several such arguments have, for example, been based on the success of science: how can present scientific theories account for such a wide range of phenomena if they are not, at any rate 'approximately', true? Van Fraassen counters by pointing out, first, that we can't ask for explanations of everything on pain of infinite regress; and, secondly, that anyway alternative explanations of science's success are possible. The explanation he himself favours is 'Darwinian':

---

[1] There seems to be no particularly good reason for the choice of this name (for van Fraassen's remarks on the name see pp. 4–5): in particular it should not be taken as suggesting that van Fraassen's position is at all constructivist in the sense made familiar by recent debates in the philosophy of mathematics.

[2] This would not, however, be van Fraassen's own way of expressing this thesis—he prefers a 'semantic' formulation (see section 2).

[3] What exactly this term is, and what exactly it should be, taken to mean is discussed in section 3 of this review.

> . . . any scientific theory is born into a life of fierce competition, a jungle red in tooth and claw. Only the successful theories survive—the ones which *in fact* latched on to actual regularities in nature (p. 40).

Independently of the merits of van Fraassen's particular alternative, his general counter-argumentative strategy is clearly bound to work. Nothing in science is going to *compel* the adoption of a realist attitude towards theories (especially given that we are allowed to regard scientists' apparent determination to hold onto certain theories despite empirical and conceptual difficulties as only being *as if* they believed their theories to be true). But this leaves open the possibility that some form of scientific realism, while strictly speaking unnecessary, is nonetheless the *most reasonable* position to adopt.

In order to explore this possibility properly, an ambiguity in van Fraassen's development of constructive empiricism must first be resolved. There is a weaker and a stronger position which he might be holding. The weaker one states that acceptance of a theory involves *at least* belief in its empirical adequacy—it is simply agnostic on the question of whether some further belief that the theory is true or approximately true is also involved. The stronger position holds that theory-acceptance involves *exactly* the belief that the theory is empirically adequate *and nothing more*.

Van Fraassen's principal argument for constructive empiricism is the minimality argument. But clearly only the weaker of the two positions just outlined is logically weaker than realism, the stronger of the two being, on the contrary, inconsistent with it. In order for the argument to entail the stronger position a further premise is required. This is that anything more than is absolutely necessary not only need not be, but *is not to be*, assumed. This extra premise is at least debatable and will generally be denied by scientific realists. In sum, the stronger position and realism are inconsistent rivals—each makes assumptions not made by the other—and surely some argument is needed in terms of the superior plausibility of one over the other, if we are to decide between them.

No such argument is supplied by van Fraassen, who, as already noted, rests content with demonstrating that particular arguments do not *establish* realism. Various traditional arguments—notably that from underdetermination—*are* mentioned, but none is systematically developed and no attempt is made to consider and rebut realist counters. Moreover, he ignores the particular kind of scientific success which has traditionally carried the most pro-realist persuasive power. This is *not* the fact that present scientific theories have many correct observational consequences. These might after all have been consciously built into the theories. The most striking sort of success has been when one theory, introduced to deal with one set of phenomena, has turned out to predict in a straightforward and uncontrived way, a completely different (and perhaps hitherto unsuspected) phenomenon. Even the most committed instumentalists, when faced with this sort of success, have found it difficult to fight off 'creeping realism'.[1] (Note that van Fraassen's own favoured evolutionary explanation of scientific success can obtain no foothold here: presently accepted theories have clearly been moulded, directly or indirectly, by the 'environment' of known empirical results, and would not be 'presently accepted' if they did not have many correct observational consequences. But this supplies no account whatsoever of why they should have successfully predicted results which were not even *in* their environment when they were first proposed.) I do not, of course, claim that this sort of predictive success *entails* scientific realism, but only that it has traditionally supplied an argument in favour of realism, an argument which stands in need of a counter from the anti-realist.

As for positive arguments in favour of definitely restricting our claims about presently accepted theories, the traditional positivistic ones are, of course, unavailable to van Fraassen. If assertions of the literal truth of theories failed to make genuine sense, then realism would have to be rejected as a matter of principle, no matter how great its superficial attraction. But van Fraassen explicitly allows that coherent sense can be made of observation-transcendent truth. If assertions were legitimate only if they could be fully justified, then straightforward claims about the (approximate) truth of our theories would have to be rejected, again as a matter of principle. But van Fraassen's account of what is involved in the acceptance of a theory requires the highly unjustifiable belief that the theory 'saves' all past, present, *and future*, all observed *and observable though actually unobserved* phenomena.

This is all on the (probably correct) assumption that van Fraassen's is the stronger of the two positions outlined above. But what if only the weaker is advocated? This weaker position says, remember, that in accepting a theory we commit ourselves to *at least* a belief in its empirical adequacy. Hence it is not inconsistent with realism. The minimality argument would, if accepted, establish this weaker position but would *not* thence supply a further argument against taking the step into some form of realism. So long as the realist acknowledged that he makes assumptions which are not strictly speaking necessitated by any aspect of science then he could accept the argument without its affecting his position.

---

[1] A notable example of a committed instrumentalist who succumbed to a degree of 'creeping realism' is Pierre Duhem. He explicitly denied that high-level theories are true-or-false, holding instead that they are classification schemes. Yet faced with some theory's prophetic success he felt forced to concede that this indicates that the theory supplies something approaching a 'natural classification'. It is not easy to see too many significant differences between a 'natural classification' and a true theory. (See Duhem [1906] and, for analysis, my [1982].)

Is there nonetheless an argument in favour of making as few assumptions as possible (and maintaining an uncommitted attitude towards the rest)? Van Fraassen seems to believe that this is almost self-evident; that, at any rate in philosophy, fewer assumptions are automatically better. But this cannot be right. After all, to take an analogy with *physical* realism, I know that in order to make sense of my sense perceptions I am not *compelled* to assume the existence of a real, external world; nonetheless, physical realism seems not only a reasonable position to take, but the only reasonable position to take.

Once the positivist line has been broken, as van Fraassen allows, what real argument is there for only going a little beyond it? There is, so far as I can tell, just one remark in the book addressed to this question. This is that no one is actually compelled to hang for a sheep rather than a lamb (p. 72). But the point of the old maxim is surely that if a little and a lot are both available, and if the penalty is in either case the same, then one may as well go the whole hog and take the whole sheep. It is not that anyone is compelled to take the sheep, but that it seems rather perverse not to. If scientists when they accept a theory are to be considered as holding beliefs (that the theory saves *all* the phenomena), which in the nature of the case cannot be justified, then why debar them from a little extra belief (that the theory is at any rate our present best guess as to the truth),[1] which they anyway seem generally to hold, and which now fails to clash with any general principle?

There is a well-known anti-realist answer to this question, but it is not clear that van Fraassen can avail himself of it. This answer is, roughly speaking, that the belief that our present theories are true is strongly at odds with the history of science. If theories are interpreted literally, then this history can only be seen as punctuated by radical revolutions: theories which were once widely regarded as true are now regarded as unambiguously false (though still of course containing many correct observational consequences). If we are not to adopt the absurdly presumptuous view that history consists of a series of false starts and that we have *now* definitely hit on the truth, then we must certainly modify and weaken the typical realist claim about the status of our theories. They can, at best, be regarded as only

---

[1] As this remark indicates, it seems to me that van Fraassen requires the realist to overstrain his belief faculties when he accepts a theory. In the sort of 'conjectural realism' which I advocate in my [1982] there is no commitment either to the truth, 'approximate truth' or even probable truth to degree x of presently accepted scientific theories. Indeed the 'conjectural realist' may well believe that they are all very probably false. This does not prevent him believing that they are true-or-false attempted descriptions of reality, both of observable reality and of the reality 'hidden behind the phenomena,' (I take this to be the *core* of scientific realism) and it does not prevent him believing that these theories constitute our presently best guesses as to the nature of reality.

'approximately' true, or as 'reflecting *to some extent* the real structure of the universe'. These notions have notoriously proved resistant to precise analysis. More fundamentally, it is not clear that, so far as high-level theories are concerned, this approach holds out any real hope of success. Is any account of the notion likely to make 'heat is a fluid' 'approximately true' (assuming that heat really consists of molecular motion)? Is it now regarded as 'approximately true' that the ether exists? On the other hand, later developments are not at odds with the idea that, *intuitively speaking*, the caloric theory and the classical ether theory were to some extent 'empirically adequate'. Even here there are well-known difficulties: superseded theories are generally inconsistent with their superseding rival even at the empirical level. Nonetheless, there remains a strong intuition that scientific development has been essentially continuous at the empirical level. If the anti-realist could produce a precise account which captured this intuition, then the resulting picture of high-level theoretical discontinuity and low-level observational continuity would seem to supply him with a strong argument against taking a realist attitude towards theories.[1] Can van Fraassen take advantage of this opportunity? This is not entirely clear because of certain difficulties with his precise notion of 'empirical adequacy'. As we shall see (in section 3), this is a very weak notion. Nonetheless if we take fully-fledged versions of the caloric theory, classical physics or whatever, *complete with all the necessary auxiliaries*, then these are inconsistent with accepted observation statements and this in turn certainly makes them empirically inadequate in van Fraassen's sense. *These* theories have been rejected as empirically inadequate no less definitely than they have been rejected as false on realist accounts. If, on the other hand, we take just the 'core' of the caloric theory or of classical physics then these do remain 'empirically adequate' on van Fraassen's characterisation—but only because, as Duhem pointed out, they make, in isolation, no claim which is directly testable at the empirical level. In this sense *any* theory is 'empirically adequate'. But what is really needed in order to solve this problem of giving a more precise characterisation of the claim that science has been 'essentially' continuous at the empirical level despite high-level discontinuities is some precise account of how once accepted but now superseded theories could, in their fully-fledged versions, have had observational consequences which, although strictly false, were nonetheless 'nearly' right. Van Fraassen provides no machinery with which to approach this problem.

To sum up this section: my claim is that if van Fraassen is advocating a genuine *alternative or rival* to realism, then he should argue more persuasively for his alternative's greater plausibility; if he is advocating merely a *weakening* of realism, then he should tell us why, rather than merely assume that, weakening a philosophical position makes it better.

---

[1] I do not believe that this argument would in fact destroy realism though it would strongly favour a move to 'conjectural realism' (see preceding footnote).

2 SYNTAX OR SEMANTICS ?

Van Fraassen characterises all his more precise methodological notions in terms of the 'semantic approach' to theories. According to this approach, theories are best seen, at any rate 'in the first instance', as consisting of sets of models rather than sets of sentences:

> The syntactic picture of a theory identifies it with a body of theorems, stated in one particular language chosen for the expression of that theory. This should be contrasted with the alternative [which van Fraassen advocates] of presenting a theory in the first instance by identifying a class of structures as its models. In this second semantic, approach the language used to express the theory is neither basic nor unique; the same class of structures could well be described in different ways. . . . The models occupy centre stage (p. 44).

Newtonian particle mechanics, for instance, should be seen as first and foremost a set of Newtonian particle *systems*, and *not* as a set of axioms.

It is not clear to me why it should matter very much how a theory is regarded 'in the first instance'. Certainly, and despite the impression which van Fraassen seems to try to convey, the primacy of the semantic approach cannot rest on logical considerations. So far as logic is concerned, syntax and semantics go hand-in-hand—to every consistent set of first-order sentences there corresponds a non-empty set of models, and to every normal ('elementary') set of models there corresponds a consistent set of first-order sentences.[1] As van Fraassen himself made clear in an earlier (and perhaps clearer) account of the matter:

> There are natural interrelations between the two [syntactic and semantic] approaches: an axiomatic theory may be characterised by the class of interpretations which satisfy it, and an interpretation may be characterised by the set of sentences which it satisfies; though in neither case is the characterisation unique. These interrelations make implausible any claim of philosophical superiority for either approach.[2]

Whether intentionally or not, van Fraassen seems to go back on this in the present work by suggesting that the semantic approach can indeed be shown to be superior on grounds of philosophical principle. For example, he argues that there are certain basic defects in the syntactic approach which are avoided by the semantic. These arguments, however, go through only against strawmen (or, perhaps, against really held but long superseded views). For instance, he produces one particular syntactic characterisation of the empirical content of a theory (set of the theory's consequences expressible in the observational sub-vocabulary); shows that this characterisation is unacceptable and encourages the inference that this means that the syntactic approach cannot at all adequately characterise the notion. More generally he criticises various particular assumptions which particular alleged proponents of the syntactic approach have made, and concludes from the success of these criticisms that the whole approach is in trouble.[1]

---

[1] The interrelationships admittedly become much more complicated if we go beyond first-order theories. But if van Fraassen wants to found his case on the insufficiency of first-order logic, he should say so explicitly and should argue explicitly why anything *essential* is left out of a first-order representation. He should finally spell out why switching to the semantic approach would help in this less clear-cut area.

[2] Van Fraassen [1970], p. 326. (I am not sure what he is getting at when he says that 'in neither case is the characterisation unique'.)

Finally, van Fraassen often asserts, as in the passage quoted above (p. 71), that the syntactic view makes the notion of a theory unacceptably language-dependent. But there is no real force in this point: the sensible axiomatiser is not the prisoner of any particular language. The choice of a language in which to express a theory will be made on the grounds of suitability or convenience without any claim being made that this is the one "true" language of the theory. Take, for example, Newton's theory as formulated using Hamiltonians and alternatively using Lagrange's equations. Even if these were originally formulated in two different languages with different sets of primitives, it is easy to express the idea that they *are* simply two different formulations of the same theory—they can either be proved logically equivalent in a suitable metalanguage or we can switch to a new wider language in which both sets of original primitives become defined terms. It could be objected that if the proponent of the syntactic view has a notion of a theory *prior* to its expression in some particular precisely defined language, then he must have some more basic language in mind in which the theory is representable. Ultimately he is committed to the notion of a truly basic language—the only serious contender for this role seems to be set theory—in which all his theories are expressible given suitable definitions. This is no doubt correct but reveals no unacceptable language-dependence, and is anyway an assumption to which van Fraassen too is ultimately committed: for he too must presuppose some system in which the isomorphisms between structures that his approach requires can be stated and established.

If then we give up the idea of establishing the superiority of the semantic approach on the grounds of logical or philosophical principle, and we surely must give up that idea, then we are left only with arguments from something like greater 'naturalness', or, perhaps, greater fruitfulness. (Just as, while admitting the logical equivalence of the two formulations, we *might* argue, say, for the superiority of the Hamiltonian formulation of classical mechanics over the Lagrangian on the grounds of the former's greater naturalness or the fact that it is more fruitful since extendable in a certain way to quantum mechanics.) Those who favour the model-approach do indeed generally rest most of their case on the claim that it presents a much more faithful picture of how scientists themselves regard theories. I doubt that this is correct, but assume that it is. Does it follow that the semantic view is a superior view of theories? Consider an analogy. First-order logic can be developed using *Principia Mathematica*-style formalisations with logical axioms and a small set of rules of proof (generally just two) or, alternatively, it can be developed using Gentzen-style so-called 'natural deduction'—which involves no axioms but a larger set of rules. As the name indicates, some people at least would argue that the Gentzen system allows a 'more faithful' picture of how mathematicians, for example, actually reason. While certainly not indisputable, there is surely a stronger *prima facie* case for this claim than for the analogous claim about the semantic view of theories. But anyway say that we accept it. There is no reason why this in itself should be taken as establishing the superiority of the natural deduction approach—for logic is out to capture correct reasoning only in the sense of providing the 'right' demarcation between valid and invalid inferences. Except for pedagogic or other pragmatic purposes, the extent to which the ways it supplies for deciding validity reflect the ways people go about making this decision does not matter at all. Moreover, there are other merits that a system of logic can have—principally, of course, ease of performing meta-proofs. Many important meta-theorems are more easily achieved for *Principia Mathematica* systems than for natural deduction systems (because of the former's relative paucity of primitive machinery). The only reasonable position seems to be to exploit both approaches and not to regard them as competitors that we must decide between. Similarly in the case of the syntactic and semantic views of scientific theories: despite van Fraassen's propaganda in favour of the former, these are complementary not rivals; there is simply no point in arguing in favour of one and against the other.

---

[1] For instance he makes a good deal of the undoubted fact that Carnap's distinction between 'meaning postulates' and 'empirical postulates' often seems arbitrary. But many philosophers whom van Fraassen would presumably classify as followers of the syntactic approach (Lakatos and Glymour immediately spring to mind) have already argued that the right reaction to Duhem's point about the logical structure of tests is *not* to try to restore direct testability of 'central' theories by lumping all the auxiliaries together as specifying the 'meaning' of the central theories' theoretical terms, but instead to allow the auxiliaries as extra empirical assumptions which may themselves be amended because of empirical difficulties. Hence Carnap's distinction can obviously be dropped without dropping the 'syntactic' view.

3 WHAT DOES IT TAKE TO 'SAVE THE PHENOMENA'?

One point of detail in the book should be commented on since it is likely to lead to confusion. One would expect someone who claimed that scientific theories aim at empirical adequacy to characterise a theory as empirically adequate when it *entailed* some given range of phenomena—or, rather more precisely, when it entailed the phenomena, *given* accepted statements of initial conditions and perhaps accepted auxiliaries. Only then could a theory be said really 'to save the phenomena'—mere *consistency* with the phenomena is not enough to 'save' them.

Most of van Fraassen's general remarks, I believe, presuppose this characterisation. But surprisingly, when he gives precise characterisations of adequacy he requires *only* consistency with the data: an empirically adequate theory is one that has '*at least one* model that all the actual phenomena fit inside' (p. 12; italics added). This seems so clearly unacceptable that one might suspect a simple slip has been made. But not so: the characterisation is repeated and some of its seemingly unacceptable consequences explicitly drawn and embraced. The most obvious such consequence is that 'empirical adequacy, like truth, is "preserved under watering down"' (p. 67). In other words, any theory logically weaker than an empirically adequate theory must itself be empirically adequate, and so, in particular, logical truths are empirically adequate.

Van Fraassen is, of course, right that truth is also 'preserved under watering down'. But this only shows that a realist must have science aim at something more than truth—namely truth plus high informative content, or, if you like the 'whole truth'.[1] Similarly the anti-realist must surely make some greater demand of his theories than mere consistency with the data. I suspect that van Fraassen has been overinfluenced and misled by the old Duhem point that what we standardly call scientific theories do not have observational consequences of their own. His favourite examples—classical and quantum mechanics—are clear cases in point. Auxiliary assumptions about particular systems must be added if either theory is to 'yield the data'. This does not, however, vindicate van Fraassen's characterisation. Even if we take these sorts of theories without auxiliaries we surely must demand more of them than mere consistency with the phenomena. It is not enough that *some* further assumptions consistent with the theory together with it yield the data. Rather the theory must yield the data when added to *particular* auxiliaries—namely those which are themselves 'correct'.

As it stands, van Fraassen's notion of empirical adequacy is itself inadequate.

4 IS CAUSAL EXPLANATION 'CONTEXT-DEPENDENT'?

Van Fraassen arrives at the topic of explanation *via* the following route. A possible argument for realism is from the 'explanatory power' of theories. Theories explain certain phenomena. In order to have such explanatory power, theories need to possess something more than mere empirical adequacy—maybe 'truth *plus*', but at any rate truth. His counterthesis is that there is indeed something more to explanatory power than empirical adequacy, but this extra factor has nothing to do with truth and is instead 'pragmatic' in character.

His argument in favour of the 'pragmatic' or 'context-dependent' nature of explanation is independently interesting and has certainly been well received. (Indeed I have heard it described as 'definitive'.) The crucial part of this argument is the attempt to demonstrate that the 'pragmatic' account solves certain difficulties which are generally regarded as intractable on the 'standard' Hempel-Oppenheim view.

I shall try to outline van Fraassen's account of explanation, then indicate why it seems to me both unnecessary and unacceptable and, finally argue that, contrary to his own explicit claims, it does not in fact solve the difficulties which afflict the Hempelian account.

---

[1] On this point see, for example, Popper [1972].

Requests for an explanation of an event can be regarded in one of two different ways: either as requests for the *cause* of an event, or as requests for an answer to a '*why-question*'. Either notion, when properly analysed, is, claims van Fraassen, context-dependent. I shall concentrate here just on the notion of cause.

The standard analysis is that if A caused B, then A was a sufficient condition for B.[1] According to van Fraassen this cannot be correct since, for example, we would certainly allow that a particular plant which died when sprayed with defoliant died *because* it was sprayed, even though we knew that the defoliant is only 90 per cent effective.

Now this is certainly no knockdown refutation of the standard analysis, for a defender of that analysis could react to the alleged counterexample as follows. 'I readily admit that my account does not directly mirror ordinary usage, but would argue that it is none the worse for that. We do indeed often speak of 'the cause' of an event when the factor concerned was less than a fully sufficient condition for the event. But this is loose (though generally perfectly harmless) talk: we could always in such situations, find extra conditions which operated and which together with the cited factor, do indeed amount to a sufficient condition for the event. The spraying with defoliant was not, strictly speaking, the cause of the plant's death, though it was an important *causal factor*. The real cause was the spraying plus certain features of the particular plant's constitution. (Of course, the fact that *we* may have no interest in finding out what these features were—because for practical purposes we are only interested in the effects of the spraying on large populations of plants and not on any particular plant—is quite irrelevant to the question of whether they were in fact part of the cause.)'

Van Fraassen must regard any such response as insufficient, since he regards the plant example as refuting the sufficient condition analysis, and sets out to produce an alternative which captures ordinary usage more directly. He first tries the idea that the cause of the event is its *conditio sine qua non*: had the plant not been sprayed it would not have died. The straightforward logical translation of this suggestion is as the thesis that a cause is a necessary condition. This, however, makes the account clearly unacceptable. For one thing, the plant obviously *might* have died in some other way. Moreover, the plant's having been put in the ground in the first place was clearly necessary for its death, but no one would say that the planting caused the death. Sticking to the idea of cause as *some sort* of necessary condition, van Fraassen is led into the slough of counterfactuals and *ceteris paribus* clauses ('had the plant not been sprayed then it would not, other things being equal, have died when it did'.) But counterfactuals and

---

[1] Notice that this analysis is not, of course, committed to the thesis that if A is a sufficient condition for B then A caused B. (This thesis is indeed clearly wrong—see *below*.) No doubt a full analysis of the notion of cause requires several extra, and rather subtle conditions.

ceteris paribus clauses require a 'pragmatic' or 'context-dependent' analysis. In fact, claims van Fraassen, the right conclusion to draw is that there is no scientific objective notion of the cause of an event. Science, which is itself *not* context-dependent, aims to describe the whole 'causal net' of necessary conditions in which an event is enmeshed. But 'the cause' of the event is characterisable only in terms of (varying) human interests: what *we* regard as *the* cause of an event, and hence what *we* regard as explanatory, are certain 'salient features' of the 'lines' within the 'net' leading up to the event. Since what appears 'salient' in one situation, and to one person, may not appear salient to another in another, the notions of cause and hence of causal explanation are context-dependent.

In the case of causal explanation van Fraassen sees the context-dependency as characterisable by two factors: the 'contrast class' (we are always in context explaining why x occurred *rather than* $y_1, y_2, \ldots\ldots, y_n$) and the 'relevance relation' (the explanandum must not only pick out x above $y_1, \ldots\ldots, y_n$, but must also bear the right relevance relation to x). Detailed criticisms could be made of this account, but more importantly I remain unconvinced that this sort of approach is, quite generally, the right way to go. The 'sufficient condition' view provides a simple analysis which admittedly is somewhat distant from ordinary usage—though explanations (in terms of hidden extra factors) for the divergences between the ideal analysis and ordinary usage are readily forthcoming. If instead we insist on reflecting ordinary usage more fully, the whole analysis becomes terribly complicated and, so far as I can see, nothing is gained. The situation is analogous to one in logic. Some philosophers argue that classical logic must be modified and elaborated because it pronounces formally invalid certain inferences which would—in 'normal' argumentative practice—be immediately accepted as valid. The counter-argument would be that we should stick with classical logic and explain this situation by showing that there are certain implicit premises which are being tacitly accepted, and which, if articulated and added as explicit premises, would render the initially classically invalid inference classically valid. The alternative, of trying to capture the validity of *some* of these inferences directly, would in fact write some of these tacit premises into the rules of reasoning themselves. This would in turn make logic context-dependent, since what is accepted as 'obvious' (and hence what can function successfully as a tacit assumption) will in general differ from context to context. While this is, of course, a possible way to go, it does seem to lead to unnecessary complexity, without freeing logic from all idealisation.

Moreover, just as in the logic case there are inferences which are *absolutely* valid (because no tacit premises are needed), so in the case of causality there are surely *absolute causes*—where any context-dependence collapses, and sufficiency is restored. This is precisely what happens in the case of deterministic theories in science. Assuming, for the sake of an example, that the 19th Century account is true, the cause of Jupiter's having position $\mathbf{r}$ at time t is it's having had position $\mathbf{r}'$ at time t-Δt and it's being a body of a certain mass in a certain force-field. The cause is, given accepted laws, a fully sufficient condition for the event. *Ceteris paribus* assumptions need not be separately made—they are, if you like, written into the cause, in particular into the claim about the *total* force acting on Jupiter. Similarly, again assuming 19th Century science, the cause of the light and dark fringes in (some particular instantiation of) the two-slit experiment is the coherence of the unobstructed light waves emerging from the two slits and affecting the screen. These conditions are, given the classical wave theory of light, *sufficient* for the production of the fringes. Taking these as the clear-cut cases—as has traditionally been done—we obtain a clear-cut analysis and lose nothing since the admitted fact that ordinary people often talk of causes which do not immediately fit this model can be otherwise explained than by giving up the model.

Finally, there is a further reason for rejecting van Fraassen's account of cause and of causal explanation. As it stands, no restrictions are imposed on 'relevance relations' and so any group seems to be free to set itself up complete with its own relevance relation and simply deny that presently accepted causes or presently accepted explanations are causes or explanations at all. Creationists, for example, can merely insist that the factors cited by neo-Darwinians to explain certain evolutionary developments fail to bear the right relevance relation to the events concerned.

I would claim, therefore, that van Fraassen has certainly not demonstrated that we *need* to incorporate explicit reference to the context in our account of cause and of causal explanation. I would also claim that there are certain dangers in doing so, which he has done nothing to lessen. But he also argues that incorporation of the context *helps* in certain ways—in particular that it leads to a solution of certain basic problems which have proved stumbling blocks to the orthodox approach, especially in the case of explanation. The sharpest of these problems come in the form of apparent counterexamples to the Hempel account: cases of intuitive explanations which fail to fit Hempel's conditions, or cases which fit Hempel's conditions but which fail intuitively to be explanations.

Perhaps the most famous counterexample is due to Bromberger.[1] The laws of geometrical optics, together with statements about the elevation of the sun and the lie of the land at the relevant point of the earth's surface, entail that a flagpole situated there is of a certain height *if, and only if*, its shadow is of a certain length. It seems to follow from Hempel's account that we can *either* explain the length of the shadow on the strength of the laws plus initial conditions, including the height of the flagpole, *or* explain the height of the flagpole on the strength of laws plus initial conditions, including the length of the shadow. Yet intuition rebels at the latter 'explanation'. Since this is the only type of counterexample for which van Fraassen develops any sort of detailed demonstration that his approach solves the problem, I shall concentrate on it here.

---

[1] See his 'Why-Questions' in Colodny (*ed.*): [1966]. Other counterexamples, concerned with 'relevance' and with low-probability 'explanations' have been highlighted by Wesley Salmon (see Salmon *et al.* [1972]). Van Fraassen does not systematically apply his account to these more challenging, counterexamples. The *impression* he creates that the problems they pose fall to his approach (which certainly includes consideration of relevance and of probabilities) is, I think, merely an impression, but I shall not attempt to argue this thesis here.

The 'relevance relation' is what is meant to supply the solution. It may be that laws plus initial conditions imply that *A* will be the case if, and only if, *B* is, and yet *A* describes one sort of factor and *B* another; when someone asks for an explanation he is, in context, asking for a factor of a particular type; hence he may be satisfied with 'Because *B*' as an answer to 'Why *A*?' (*B* bears the right 'relevance relation' to *A*), and yet not accept 'Because *A*' as an answer of 'Why *B*?' (the relevance relation need not be symmetric). In the flagpole example, we are usually, in context, asking for a 'mode of production' (p. 131). Hence since the flagpole produces the shadow and not *vice versa*, the flagpole's height, in such contexts, may explain the shadow's length, but not *vice versa*.

One point of detail is that a 'mode of production' is surely a cause by any other name. Hence it ought, on van Fraassen's account, itself to be 'context-dependent' and therefore not employable in a specification of context. But still it is of course true that we think flagpoles cause shadows to have a certain length, but do not think that shadows cause flagpoles to have a certain height. I would claim that this so-called contextual factor is part and parcel of the notion of a causal explanation: in deterministic science we are *always* looking for causes or 'modes of production'. If this is right then the problem posed by the flagpole example is just that it shows that a condition may be sufficient, in certain circumstances, for the occurrence of an event and yet not be a cause of this event: there is an extra intuitive, or metaphysical component to our conception of cause which has eluded the formal analysis; but there is certainly no case in which the length of a shadow explains the height of a flagpole. If, on the contrary, van Fraassen is right then we would at any rate *expect* there to be such cases—cases in which the relevance relation is different and where shadow lengths do indeed bear the right such relation to flagpole heights. Van Fraassen claims to have produced precisely such a context.

He tells an X-certificate story with himself as a hero. On a visit to the estate of an old French chevalier, he is taking a pleasant tea with his host on the terrace when the shadow of a tower built in the grounds falls on the terrace (dramatic licence having turned the flagpole into a tower). The shadow makes it too cold for tea to continue comfortably. Exclaiming 'Why does the tower have to have such a long shadow?' our hero is taken literally and is told that the tower commemorates a visit of Queen Marie Antoinette to the estate, was built on the spot where the chevalier's ancestor first greeted the queen, and is the same height in feet as the age the queen would have achieved in the year that the tower was built. Later that evening, however, he is told a quite different story by a mysterious chambermaid: that the chevalier had fallen in love with a maid whom he subsequently murdered; that the tower marks the spot where he murdered her; that he had vowed that the shadow of the tower would with every setting sun cover the terrace where he had first declared his love; and that he had chosen the height of the tower accordingly.

This version has been bowdlerised for publication in this family magazine. The full story can be found in van Fraassen's book. What *cannot* be found there, however, is a case of an explanation of the height of a tower by the length of its shadow. Unless, that is, we are allowed to cheat quite blatantly by changing the event to be explained to suit our purposes. The problem arose remember within physics—geometrical optics yields a biconditional in a circumstance in which the causal arrow definitely flies only one way. If we switch to the realm of human decisions, then, of course, the problem may well disappear. Someone may decide to bring about x directly, or because physics, say, tells him that x will have effect y and y is what he really wants to achieve. But to say in the latter context that *y caused x* is simply a confusion. If you like, it was the agents' *desire to achieve y* which caused *him to bring about x*. Van Fraassen's story only illustrates this obvious lesson—it provides two competing (*not* complementary) explanations of *the chevalier's decision* to build the tower to a certain height. It is therefore completely independent of the problem at hand. This is shown by the fact that van Fraassen's trick could just as well be pulled in cases where the causal situation and the logical formalism of our theory go hand in hand—both being asymmetric. Suppose some scientific theory yields the conditional $A \rightarrow B$, does *not* yield the conditional $B \rightarrow A$, and that we are quite happy to say that A explains the occurrence of B (the necessary theory being assumed), and finally that there is no question of B similarly explaining A. Except, that is, on van Fraassen's account. For suppose that a teacher has chosen to realise circumstances A because he wanted to show his students effect B. Then if we are allowed to say that we can explain the height of the tower by the length of its shadow (the chevalier wanted something with a shadow that long), then we should equally well be allowed to say that we can explain the cause A in terms of its effect B (the teacher wanted some cause that would have effect B). This is absurd.

Van Fraassen's treatment does nothing to solve the problem posed by Bromberger's example. He has given us no reason to import the 'context' into our formal account of scientific explanation.

JOHN WORRALL
London School of Economics

REFERENCES

COLODNY, R. G. (*ed.*) [1966]: *Mind and Cosmos*. University of Pittsburgh Press.
DUHEM, P. [1906]: *The Aim and Structure of Physical Theory*. Reprinted 1962 by Atheneum.
POPPER, K. R. [1972]: *Objective Knowledge*. Oxford University Press.

SALMON, W. *et al.* [1971]: *Statistical Explanation and Statistical Relevance.* University of Pittsburgh Press.
VAN FRAASSEN, B. C. [1970]: 'On the Extension of Beth's Semantics of Physical Theories', *Philosophy of Science,* **37,** pp. 325–39.
WORRALL, J. [1982]: 'Scientific Realism and Scientific Change', *The Philosophical Quarterly,* **32,** pp. 201–31.

Brit. J. Phil. Sci. 35 (1984), 81–100 *Printed in Great Britain*

Reviews

JENSEN, V. J., and HARRÉ, R. (eds) [1981]: *The Philosophy of Evolution*, Harvester Press. Pp. vii + 299.

The idea behind this book, and the seminar and symposium that helped produce it, was that the time had come to look at the use of evolutionary notions in different areas from the philosophical point of view. The result is a set of papers on different topics, and an almost ingenious thread of commentary, particularly from Rom Harré, pointing up relationships to the declared common denominator.

Although the idea of sustained unidirectional change over a sequence of generations wasn't an invention of biology, biology it is that has founded upon it so elaborate and finely wrought a structure as to make it seem today biological in its essence. And indeed most of the present volume is directed at the proposition that the understanding of human affairs may benefit from the application of *biological* ideas.

Why should we think of looking to the evolutionary concepts of biology for help in trying to understand human life and social change?

One answer is that theories of biological change may be helpful because the sort of change they account for is structurally parallel to the sort of change that concerns the student of human history: in particular, directional change that is manifested through the passing of generations. Some in this context debate the relevance of evolutionary concepts to social history—Harré's paper, *e.g.*, which tries to assess the degree of analogy between on the one hand the relationship between species, organisms and genes and on the other cultures, institutions and rules. Dōbert argues for the legitimacy of analysing social evolution into a succession of stages, and makes his point by presenting in some detail his highly interesting interpretation of the European witch craze.

Others focus upon a particular aspect of human history and discuss the relevance of biological ideas there: for example, Mitelstrasse, Laudan and Newton-Smith debate the application of these ideas in the explanation of the history of science. Toulmin explores the parallels by distinguishing some different senses of *adaption*.

Though Marxism offers probably the most influential account of social change, the Marxist contribution to the present volume is a disappointingly negative one. Both Ruben and Keiter take their stand on the view that it is the activity of production which distinguishes humanity and that the rôle of biological theories in understanding human history must therefore be negligible. What I miss in these papers, however, is much real contact with either anthropology, to substantiate claims which otherwise remain interesting (or, as Harré claims, mistaken) speculation, or the biological ideas which are after all in question, the possibility of their relevance is ruled out from a very lofty seat of judgement.

A bolder answer to the question of the relevance of biological evolution argues that the theory which explains the latter necessarily applies to other change having the same essential features: and that these features are indeed possessed by some of the processes of change that occur in the human sphere. The object then is to articulate the developed theory of variation and selection in properly general terms, so that not only will the evolution of species be explained, but perhaps the 'evolution' of ideas and the 'evolution' of societies also.

Hull reminds us in his paper that two conditions have to obtain before any process of 'evolution by selection' can occur: there have to be entities capable of producing copies of themselves, and there has to be some consistent interaction between those entities and their environment such that the survival of each entity is influenced by its characteristics. Dawkins has introduced the term *replicator* to express these points succinctly: a *replicator* is any entity which produces copies of itself; and evolution occurs if the environment is such that the characteristics of replicators have a consistent influence on their own survival.

Hull argues that the conceptual situation will be easier if we introduce a further category of entity to distinguish this second desideratum specifically: '*interactors*—entities which produce differential replication by means of directly interacting as cohesive wholes with their environments' (p. 33). Much of the controversy over the units of evolution is thus, he claims, to be dissolved away.

For example, those who insist on the gene as the unit of evolution—like Dawkins—are then seen as asserting that genes alone are *replicators*; while those who ascribe the key role to the organism are construed as asserting that organisms are the important *interactors* in organic evolution. The two positions then appear complimentary rather than incompatible.

As it stands, Hull's definition of *interactor* is rather too general for his purpose. *Anything* in an environment which contained replicators might affect which of them survived, and so would count as an interactor, on the definition offered. What Hull wants, rather, is to catch in general terms the relationship between an anthill and its ants: between the genes, and the body they both produce and 'inhabit'. It is *this* relationship—which also obtains at a higher level between organisms and social structures—which gives occasion for disagreements over the 'unit' of evolution, and in order to generalize the theories of biological evolution one wants the notion of a replicator which is not only a vehicle for replicators, but also the product of its passengers.

The fact that the subjects of biological evolution embody these highly unusual relationships suggests that writers like Hull may look in vain for concepts and theories which cover both biological and, for example, intellectual change. But the attempt is a stimulating one and provides I think the most interesting contribution of the present volume.

VERNON PRATT
Lancaster University

NOZICK, R. [1978]: *Philosophical Explanations*. Oxford University Press. £15. Pp. xii + 764.

This book raises high hopes. Written by a philosopher of established reputation, in what one would expect to be the prime of his life, and ranging in its 700 pages over so many of the grand old themes—personal identity, explanation, knowledge & scepticism, free will and punishment, foundations of ethics, the meaning of life, the nature of philosophy itself—one opens its pages with appetite whetted by the hyperbole of the publishers and of some early reviewers—"Nozick presents a new mode of philosophizing. . . . With *Philosophical Explanations*, he emerges as a major philosopher of the twentieth century." "It will change the way philosophy is done."

This reviewer can only report that his hopes were largely disappointed, and that his appetite was soon sated by seemingly endless quantities of superficially dazzling philosophical prose, from which however he has digested rather little of nutritive value for his intellect. If this book does change the way philosophy is done, it will, I fear, be for the worse. Much of what seems new in its mode of philosophising looks more like a throwback to the style of windy, undisciplined metaphysical speculation that one associates with the nineteenth century. I am prepared to be excited by an author who can refer to Aurobindo, Buber, Fichte, Frye, Hegel, Heidegger, Nietzsche and Picasso, with apparently equal facility as to Carnap, Einstein, Gödel, Goodman, Kripke and Quine. But although Nozick's intellectual intake is clearly gargantuan, I am not so sure about the quality of the thinking which it has nourished.

In the first chapter, The Identity of the Self, Nozick considers questions of personal identity in terms of a 'closest continuer' schema for identity of temporal entities generally, according to which something at t_2 is not the same entity as x at t_1 unless it is x's closest continuer (on some dimension of closeness appropriate to the kind of entity concerned). I find this schema highly contentious—how can we identify some *thing* at an instant t_1 (or t_2) without *already* using a criterion of identity for that thing over its temporal duration? But I can hardly embark on detailed criticism here. Much though Nozick has read, he shows no awareness of Wiggins' work on identity (*Sameness and Substance*), so although the first half of this chapter is undoubtedly worth reading, it can hardly be regarded as a definitive contribution to the subject. The second half marks, in my view, a quite unnecessary regress to metaphysical speculation, in Fichte's style, about how the self synthesises itself. Surely the real question is under what conditions should we say that a creature is conscious of itself, and capable of referring to itself?

In Chapter Two, Nozick shows courage (or foolhardiness?) in taking seriously the question 'Why Is There Something Rather Than Nothing?'. And in allowing a diagram of 'Nothingness Force' to be published, he risks bringing philosophy into even more disrepute than it already suffers. There may well be interesting questions to ask about ultimate explanation, and about mystical experience, but Nozick did not persuade me to take this quite extraordinary chapter seriously.

Reviewers have already picked out the third chapter, on Knowledge & Scepticism, as the one jewel in a bag of stones, and I can see why. In it, Nozick presents a novel analysis of the concept of knowledge, according to which the conditions which must be added to S's truly believing that p, to make that belief *knowledge* that p, are the 'tracking' conditionals:

If p weren't true, S wouldn't believe p
If p were true, S would believe p.

Nozick then applies this analysis to answer the challenge of scepticism, and offers, not a proof that the sceptic is wrong, but an explanation of how knowledge is possible even if (as the sceptic alleges) it is logically possible that we are dreaming, or are disembodied brains floating in a tank.

In the Introduction to the book, Nozick generalises this to a methodology for philosophy generally, according to which one may simply show how what one claims is possible (*e.g.* by explaining how it is after all compatible with something that appears inconsistent with it), rather than 'coercively' trying to make someone else believe one's claim (by forcing a conclusive proof on his attention), which, Nozick suggests, 'is not a nice way to treat anyone'. The title of the book is meant to encapsulate this supposedly 'non-coercive' mode of philosophising.

Whether Nozick's analysis of knowledge is correct, and whether it makes possible the defence against scepticism which he recommends, is already the subject of critical attention in the journals (see Graeme Forbes, forthcoming in *The Philosophical Quarterly*, and Crispin Wright, forthcoming in *Analysis*). Clearly, much will depend on the interpretation of those crucial conditionals, a matter to which Nozick devotes rather little attention.

I doubt whether the superficially attractive distinction between 'non-coercive' and 'coercive' modes of philosophising really amounts to anything. For even the former, insofar as it is to achieve any objective results at all, must appeal to logical relations such as compatibility (*i.e.* lack of entailment of the negation), but this is no more or less 'coercive' than the appeal to entailment in traditional proof. Surely the opprobrious term 'coercion' is better reserved for ways of inducing someone to believe something *not* by offering reasons (which the offerer may hold to be logically conclusive), but by promises, threats, deprivation, drugs, or torture. He who presents arguments in an effort to persuade is not thereby a brainwasher; and woe betide us if we ever forget that distinction.

The fourth chapter, on Freewill, opens the section of the book entitled 'Value' which comprises more than half its length. Nozick makes valiant efforts to provide an account of rational or moral action tracking value in the way he has claimed that belief tracks the facts when it constitutes knowledge. So the idea from which he starts is that S's doing A 'tracks rightness' when act A is right, and S intentionally does A, and:

If A weren't right, S wouldn't intentionally do A,
If A were right, S would intentionally do A.

Nozick soon finds, however, that no such theory illuminates his worries about freewill, and, as he warns us, this chapter contains 'more thrashing about' than any other. Yet he appears to consider even the most inconclusive of his lucubrations worthy of the immortality of OUP's print, and of the time of such readers as have the patience to follow his extended meanderings.

There follows an extremely lengthy chapter (170 pages—a book within a book!) on Foundations of Ethics, on which I hardly feel qualified to comment, except that in his earnest effort to 'explain how there can be objective values and ethical truths', Nozick starts off very unconvincingly with the suggestion that the dimension 'degree of organic unity' seems to capture our notion of degree of intrinsic value (as applied to works of art, to living beings, and to scientific theories). By this criterion, Nozick's own book will not rank very high. However, there may well be some ideas buried in this chapter which anyone still interested in meta-ethics may value.

The sixth and last chapter is entitled 'Philosophy and the Meaning of Life'. Nozick finds that for anything to have meaning (in the sense in which we speak of the meaning of someone's life), it must transcend its limits, by connecting with things or values beyond itself. Asking for the meaning of the latter starts a regress which Nozick thinks can be stopped if we encounter a completely unlimited entity which is its own meaning. But he acknowledges that neither experience nor argument demonstrates the existence of any such entity. This inconclusiveness is rounded off by a typically fine-sounding, high-minded, but to my mind unilluminating peroration about philosophy as an art form.

I think Nozick has over-ambitiously attempted to pack into one huge book more than any one philosopher, even one as generously endowed with versatile talents as he, can hope to deal with satisfactorily. I can see that at least three smaller, better, more unified books—on identity, knowledge, and ethics—might have been salvaged and reconstructed from this overweight ramshackle structure. Was there not a candid colleague, or an acute publisher's reader, to offer such advice? Or was it offered but not taken?

LESLIE STEVENSON,
University of St Andrews

O'HEAR, ANTHONY [1980]: *Karl Popper*. London: Routledge and Kegan Paul. £9.75. Pp. xii+219.

Anthony O'Hear has given us a spirited and enjoyable study of Popper's philosophical work, setting out the principal theses and arguments accurately and sympathetically but subjecting them to radical and pitiless criticism. The aim of the series to which the book belongs is to give 'an essentially analytic and critical account' and 'a contemporary assessment' of the work of major philosophers. In the present case this evidently includes identifying and developing the kinds of criticism which an occupant of a Popperian position—in Britain at least—can expect from his philosophical colleagues. In realising this part of the series' aim O'Hear has done his work well.

Of the book's nine chapters the first six are about Popper's general epistemology and methodology and the last three are about, respectively, probability and indeterminism, reason and society, and evolution and world 3. O'Hear argues that Popper's defence of body–mind dualism, his case for the autonomy of world 3, and his evolutionary view of perception (as signal-decoding) are all unsuccessful. On the social sciences his conclusions are that Popper has not excluded a moderate historicism working with explained trends, that—quite apart from Popper's more recent acknowledgment of the autonomous world 3 development of institutions—methodological individualism is either false or only verbally different from a moderate holism, and that the societal factors in the description of action make explanation in the social sciences importantly different from explanation in the natural sciences. On probability he argues, against Ackermann (*cf.* his [1976], pp. 71–82), that Popper's basic objection to the subjective theory stands but that his propensity theory fails. A recurring theme is the dualism of transcendent and transcended: with the Cartesian mind and world 3 as with Popper's view that there is a transcendent reality beyond our theories even though we cannot reach out beyond them to grasp it. O'Hear questions the dualism and favours trying harder to cope without postulating the transcendent.

Epistemological matters occupy the greater part of the book. O'Hear goes to the heart of Popper's theory of knowledge and takes issue with its central tenets. Popper has sought to present an account of knowledge and its growth assigning no role to inductive inferences or assumptions. Where other thinkers have taken induction to be indispensable, Popper has dispensed with it in favour of falsification, falsifiability (with precision, universality and simplicity), empirical content, truth and verisimilitude, arguing that any supposed properties of knowledge—such as reliability, certainty, well-foundedness, attainability by a method—which are not accounted for with the aid of these ideas are properties which interesting knowledge does not have. At the same time he has offered a principle for demarcating science from non-science and has taken a firm stand against relativism, instrumentalism and irrationalism. O'Hear argues that Popper's demarcation rule fails (although a generalisation of it, marking off rational activities as those in which we eliminate and learn from errors, is fruitful), that only its failure saves Popper's theory of science from sharing the main defect of instrumentalism (*viz.* its leaving the observation-remote parts of science with a no more than heuristic role), that Popper was wrong to concede the logical superiority of irrationalism to comprehensive rationalism, and that, while relativism—which is encouraged by Popper's thesis that all observation is theory-impregnated—can be defeated by a Davidsonian argument, recourse to that argument is ruled out for Popper by the thesis that all observation statements are hypothetical. O'Hear's 'general strategy in considering Popper's epistemological views is to question whether his or any other account of knowledge can do without at least provisionally accepted and partially justified certainties' (p. vii). He examines 'Popper's attempt to dispense with induction' and finds it unsuccessful since 'inductive reasoning, removed from one part of the picture, crops up in another' (pp. 57–58). He goes on to present in Chapter IV an argument to establish that 'the use of inductive policies is to be seen ... as a necessity for any conceptual scheme which provides any basis for a distinction among our experiences between the real and the imaginary' (p. 61) and, in Chapter V, an argument to establish 'that without the ability to treat some observations of enduring objects as certain and non-hypothetical we would lose the ability to fix the sense of any observation terms at all' (p. 88). The conclusion of his critique of Popper's epistemology is thus not only that Popper's actual arguments fail but further that no improved or different arguments would do the trick either.

All this is to the good. Mere exposition of Popper's contributions would serve little purpose; Popper's own expositions are plain and readable and in addition we already have Ackermann [1976] and some of the essays in Schilpp [1974]. And worthwhile criticism cannot be restricted to peripheral matters or be merely sketched in outline. Bold and developed criticism is no more than O'Hear owes to his subject and to his readers. To be sure, he finds traces of inductivism at various places in Popper's thought: in the role assigned to background knowledge in experimental testing (pp. 42–7), in the thesis that knowledge is possible only where there are regularities (pp. 62–63), in the acknowledgement of indicators of relative verisimilitude (pp. 66–67), and perhaps in the departure from a narrowly Darwinian view of the evolution of hypotheses (p. 176). Now merely to pick on isolated passages which are 'foreign to the general tenor of Popper's epistemology' (to borrow a phrase used about Popper's sincerity requirement in Musgrave [1974], p. 576) would be an unhelpful form of criticism. But that is not what O'Hear is doing. He thinks these traces are like Popper's lack of success in defending the rationality of acting on well-corroborated theories (pp. 39–42) in that they are symptoms of the hopelessness of Popper's undertaking: because there cannot be an entirely non-inductivist account of human knowledge, it is no wonder that Popper has not produced one. In taking and defending the view that these un-Popperian elements in Popper's thought are more than mere lapses from adherence to a sustainable Popperian line, O'Hear does risk giving too little prominence to Popper the proud and defiant anti-inductivist and too much to a failed anti-inductivist or unprofessing inductivist. That risk is the cost, it seems to me, of attacking Popper's theory at its heart.

If the strategy is entirely proper there are still some imperfections in its execution. Apart from occasional carelessness and obscurity in the writing, there is vacillation between talk of verisimilitude as though it were epistemological (pp. 29, 47) and talk of it as though it were not (p. 30), and there are some smallish inaccuracies ('the fact that a theory survives a severe test . . . represents an increase in its truth content', p. 31). Failure to take Popper's position seriously, while not pervading the whole book, does spoil some parts of it. Thus in supposing the adoption of instrumentalism to be a natural step from fallibilism and non-justificationism (p. 91) O'Hear tacitly assumes that conjectural theories are simply not worth having. Also he lets an important part of his argument for the non-hypothetical character of some empirical judgments depend on the claim that 'in calling something a hypothesis or theory we imply that there is some evidence which we could give for it which is more certain or obvious' (p. 80) although it is entirely foreign to the tenor of Popper's epistemology to imply any such thing. In other ways, too, his arguments are not always successful. His no-unsatisfied-standard argument (p. 150) against the irrationality of choosing rationalism can be readily defeated, or else rescued only at the cost, I think, of facilitating a parallel argument against the irrationality of choosing irrationalism; and the point that an irrationalist cannot maintain his position while pointing out its logical superiority (*ibid.*) does nothing to show that it may not *be* logically superior (much as, according to O'Hear, a predictor's future may *be* predictable even though he cannot himself produce the predictions—*cf.* pp. 144–45).

A more central example of an unsuccessful argument is the transcendental argument of Chapter IV. This is supposed to show not indeed that inductive inference must be successful but that it is indispensable for the distinction between subjective and objective. I find the argument confused and hard to follow, but its core seems to be this: if we did not have a lot of expectations of regularity which continue to be fulfilled we should not be able to distinguish, within our experience, between experience of things beyond ourselves and experience which is not of things beyond ourselves; but we are; so we do. The conclusion does not reinstate induction or threaten Popper. The argument makes a connection between believing in an external world and having expectations. 'My recognizing that a certain object is a desk leads me to expect that I will have certain experiences in the future, for example, that I will see the desk again when I look in a certain direction, that it will continue to feel hard to my touch and so on' (p. 59). But where does induction come in? At different points on page 59 O'Hear has the phrases 'on the basis of past patterns within our experience', 'based on the regularities noted within the past perceptions . . .', 'expectations built up on his past experiences' and 'inductively based expectations'. But it is whether the relevant expectations are in any inductive way based on the past that is at issue, and on this O'Hear's transcendental argument is silent. The inductivist phrases are merely ornamental. According to O'Hear the argument suggests 'that, in order to recognise that our perceptual faculties are sound and telling us how things actually are, we need to have a large proportion of our inductive expectations fulfilled' (p. 60). But it does not. It suggests rather that, in order to have continuing perceptions of external objects, we have to have expectations a lot of which are fulfilled. As far as the argument goes O'Hear has not made any 'examination of the connection between induction and belief in an external world' (p. 61). In all the talk about regularities and expectations he omits to say how *induction* is involved in the coherent experience of external objects.

There are not very many books on Popper's philosophy. The topics considered briefly but effectively by O'Hear in the second half of Chapter viii, 'Reason and Society', are treated at length in Wilkins [1978]. A large part of the latter is devoted to Popper's dualism of facts and standards, a subject not considered by O'Hear. Johansson [1975] is another book which is not as ambitious in scope as O'Hear's, since it focuses on Popper's methodological rules, articulating them all and closely criticising them one by one. Yet, while its approach is thus quite different and it contains different criticisms, it deals with many of the same topics. Like O'Hear, Johansson would like singular observation statements not always to have the hypothetical character that Popper attributes to them. However, the psychological theory of perception which he reports and apparently regards as promising fails to provide a response to the Popperian point that an object may at a later time behave in a way that renders the term 'glass', say, inapplicable to it at an earlier time; O'Hear on the other hand has offered a response to that point (pp. 79–80). In some ways—for example the Kantian themes and the inductivism at the observational level—O'Hare's book recalls Wellmer's penetratingly argued study [1967]. But the latter is concerned only with Popper's general methodology and epistemology, and not with probability, indeterminism, social science, etc., and —obviously— it takes no account of Popper's recent writings; it does have especially interesting chapters, on the empirical basis problem and Popper's decisionism, which have no clear counterpart in O'Hear's book.

The book to which O'Hear's is most similar is Ackermann [1976]. They are about the same length, both set out to be expositions and appraisals of Popper's main philosophical theses and arguments, and both achieve this and avoid serious distortions (such as the suggestion that Popper proposed falsifiability as a meaning criterion, although Ackermann confusingly uses Lakatos's term 'dogmatic falsificationism' to refer to such a proposal). Apart from minor differences of emphasis and material, what is distinctive of

O'Hear's study is its radical and sustained criticism of Popper from a position which, while it is sometimes given prominence, is never quite permitted to block our view of the book's subject. For my taste, this gives O'Hear the advantage. In any case his study is a welcome addition (despite some solecisms) to the available books on Popper.

JAMES M. BROWN
New University of Ulster

REFERENCES

ACKERMANN, R. J. [1976]: *The Philosophy of Karl Popper*. Amherst: University of Massachusetts Press.

JOHANSSON, INGVAR [1975]: *A Critique of Karl Popper's Methodology*. Stockholm: Scandinavian University Books.

MUSGRAVE, ALAN E. [1974]: 'The Objectivism of Popper's Epistemology' in Schilpp [1974].

SCHILPP, PAUL, A. (*ed.*) [1974]: *The Philosophy of Karl Popper*. Books I and II. La Salle, Illinois: Open Court.

WELLMER, ALBRECHT [1967]: *Methodologie als Erkenntnistheorie*. Frankfurt: Suhrkamp.

WILKINS, B. T. [1978]: *Has History Any Meaning? A Critique of Popper's Philosophy of History*. Brighton: Harvester Press.

VAN DALEN, D., (*ed.*) [1981]: *Brouwer's Cambridge Lectures on Intuitionism*. Cambridge University Press. Pp. 109.

L. E. J. Brouwer delivered a series of lectures at the University of Cambridge between 1946 and 1951 which he subsequently subjected to extensive revision but never published. In these lectures Brouwer exposits the fundamental concepts and theorems of intuitionist mathematics. Most of the content of these lectures appears in Brouwer's publications and can, therefore, be gleaned from the papers in the Heyting collection of Brouwer's works. The revised version of these lectures, however, consists of a systematic presentation which seems to have been intended as an introduction to Brouwer's intuitionist mathematics. This manuscript is only now being published in *Brouwer's Cambridge Lectures on Intuitionism*, edited by D. van Dalen. van Dalen has also added some explanatory remarks in the text and an appendix containing notes which Brouwer had filed with the manuscript.

Most of the material in this manuscript is probably very familiar to Brouwer scholars and students of his intuitionist mathematics. However, the significance of this manuscript seems to lie less in the material presented than in its status as an introduction to intuitionist mathematics. And since there is a great deal of interest in intuitionist mathematics among philosophers today as well as widespread confusion about its difference from, and relation to, classical mathematics, such an introduction may seem to be of some interest to the general philosophical community. My concern in this review will be with the extent to which this book constitutes an introduction to intuitionist mathematics for the rest of us—those who understand only classical mathematics.

It is well known that one of Brouwer's fundamental divergences from classical mathematics is his rejection of the law of the excluded middle. And it is clear that, while he rejected the law of the excluded middle, Brouwer defended other logical laws (see, for instance, "The Unreliability of Logical Principles" in Heyting [1975], where Brouwer defends syllogism and the principle of contradiction.) There is, consequently, a natural temptation to view Brouwer's intuitionist mathematics as what mathematics becomes when the law of the excluded middle is abandoned as a logical law. Intuitionist mathematics might seem simply to be a mathematics in which fewer inferences are licensed. This view is further supported by the fact that there is formal work which has its roots in Brouwer's mathematics, which is called "intuitionist mathematics" and which seems eminently readable and usable by classical mathematicians.

Brouwer's lectures, however, should make it clear that this view of his mathematics is mistaken. It isn't just that Brouwer's mathematics cannot be regarded as a special field of mathematics (called, perhaps, "constructive proof"). Brouwer's intuitionist mathematics is *more* than a body of formal work. And his remarks make it clear that, for the classical mathematician to come to an understanding of intuitionist mathematics (in his sense), it will be necessary for her to change her very understanding of the notion of proof. My concern in this review, then, will not be with the extent to which Brouwer's lectures constitute an introduction to recent mathematical developments, but with the extent to which they constitute an introduction to *Brouwer's* intuitionist mathematics. I will be using "intuitionist mathematics", throughout this review, as shorthand for "Brouwer's intuitionist mathematics".

The lectures begin with a section on the history of our understanding of the relation between logic and mathematics which ends with the introduction of intuitionism. In this section, Brouwer identifies, as a part of the first act of intuitionism,

> completely separating mathematics from mathematical language and hence from the phenomena of language described by theoretical logic, recognizing that intuitionist mathematics is an essentially languageless activity of the mind having its origin in the perception of a move of time (p. 4).

While the role assigned to the perception of a move of time may not be very clear, other details of this first act of intuitionism should not be all that unfamiliar to the classical mathematician. After all, mathematicians (even classical mathematicians) rarely work within axiomatic systems and the activity of coming up with a proof or construction rarely consists in logically deducing a conclusion from a set of premises. For purposes of creative mathematical work, vague pictures are often of much more use than formal systems. Thus there is a sense in which a significant part of the activity of a (classical) mathematician is languageless.

On the other hand, there is an important sense in which this first act of intuitionism constitutes a rejection of the perspective of the classical mathematician. It would be extremely odd for those of us who understand only classical mathematics to agree that this languageless activity *is* mathematics. From the point of view of someone who understands only classical mathematics, this languageless activity seems to be—not the mathematical justification of a result—but rather only a psychological process which enables the creative mathematician to come up with the result. On this view, the reasoning which is of concern to mathematics and not psychology is constituted by the justification of mathematical results. And a mathematical result is not considered to have been justified unless it has been stated and explicitly proved. The link between language and mathematics, on this view, is obvious. And there is, similarly, a link between logic and mathematics. For, once the justification of a mathematical result has been linguistically expressed, (classical) logic provides us with criteria for evaluating it. Thus, from this point of view, logic provides us with a system for evaluating *all* mathematical reasoning.

It seems clear that Brouwer means, in the first act of intuitionism, to be severing these links between logic and mathematics and language and mathematics. And it is especially instructive to examine the consequences of severing the link between logic and mathematics. Brouwer's aim is not just to reform classical logic, but to argue that logic should not be taken to play the sort of fundamental role in mathematics which classical mathematicians take it to play. At best, logical principles (whether intuitionistically acceptable or not) can provide us with nothing more than a means for evaluating linguistically expressed inferences. But, according to Brouwer, mathematics goes beyond the bounds of linguistic expressibility. Hence intuitionist logic cannot be an alternative system for evaluating all mathematical reasoning.[1] And, in robbing logic of its universal applicability, Brouwer means to rob it of its special normative role.

Although Brouwer says little in the Cambridge Lectures about what role is left for logic, he is more forthcoming on this subject elsewhere. He says, in "The Unreliability of Logical Principles", Heyting [1975], page 108,

> the function of the logical principles is . . . to describe regularities which are subsequently observed in the language of the arguments.

Brouwer's view is that the mathematical interest of logic lies in the formal structure displayed by the descriptions of these 'regularities'. He says, in

---

[1] Actually, it seems pretty clear that Brouwer would want to make the stronger claim that logic cannot play any role in evaluating inferences. See, for instance, the discussion of Hilbert in Chapter III of his thesis, and "Consciousness, Philosophy and Mathematics", both in Heyting [1975]. This stronger claim does not follow straightforwardly, however, from the first act of intuitionism.

"The Effect of Intuitionism on Classical Algebra of Logic", Heyting [1975], page 554,

> Fortunately classical algebra of logic has its merits quite apart from the question of its applicability to mathematics. Not only as a formal image of the technique of common-sensical thinking has it reached a high degree of perfection, but also in itself, as an edifice of thought, it is a thing of exceptional harmony and beauty.

Brouwer then goes on to say that the same is true of twentieth-century symbolic logic.

The views outlined above on the nature of logic and its relation to mathematics seem to be fairly accessible to those of us who understand only classical mathematics. But, unfortunately, an understanding of these views does not take us any closer to an understanding of intuitionist mathematics. Rather, it only reveals more vividly the extent of the difficulties involved in coming to understand intuitionist mathematics. The obvious first strategy for coming to an understanding of intuitionist mathematics is to work through Brouwer's proofs with the intention of identifying the criteria which determine which proofs are intuitionistically valid (and, perhaps, comparing these with the relevant criteria for classical proofs). But it should be clear, by now, that this strategy won't work. For this amounts to trying to come up with an intuitionist logic which will play the role for intuitionist mathematics that our logic plays for classical mathematics—a role which Brouwer explicitly denies that logic can play.

And it is not obvious what other means are available for the classical mathematician to come to terms with intuitionist mathematics. We can, of course, examine Brouwer's actual proofs and, in particular, how his means of expression differ from those of classical mathematicians. A cursory glance at Brouwer's lectures reveals, for instance, that his proofs are expressed much more informally than those of most classical mathematicians. But it is not at all clear that we can extract explicit, substantive, differences between intuitionist and classical mathematics from this. As van Dalen remarks, Brouwer *could* have used the formal notation "$\neg\neg\alpha$" rather than "the absurdity of the absurdity of α" (p. xi). Indeed, it seems impossible that one could extract any explicit, statable, criteria for intuitionist proof from Brouwer's language, since any attempt to do so could not go beyond the bounds of language. And this, I think, is the point. It cannot suffice to follow Brouwer's linguistic expressions of his proofs for, as he says in these lectures (p. 5)

> language plays no part other than that of an efficient, but never infallible or exact, technique for memorizing mathematical constructions and for communicating them to others.

We must, in following these linguistic proofs, somehow recognize the languageless mathematical procedure which they are supposed to evoke. And this recognition of the languageless mathematical procedure must inform our recognition of certain written proofs as representations of legitimate (or illegitimate) intuitionist proofs.

I have tried to make it clear that, however one can come to an understanding of intuitionist mathematics, it cannot be done simply by working through a series of proofs. For written proofs—whether Brouwer's or anyone else's—can only be heuristics which enable us to recognize real mathematical constructions (which cannot, themselves, be accurately described in language). Thus it is important to note that this volume is, for the most part, a book of (however informally expressed) proofs. For a philosopher who understands only classical mathematics, the central problem has to do with how we are to make sense of what Brouwer takes himself to be doing in formulating these proofs. And to use this book as an introduction to intuitionist mathematics, it is necessary to work out some means of using Brouwer's discursive introductory remarks to interpret his proofs.

It should be clear that most of the problems for understanding intuitionist mathematics which have been identified here are not problems with Brouwer's book. Although his heuristics might be more effective if a longer discursive introduction were included, it is not clear that this is so. The main point is that his proofs and definitions are no more than heuristics and, in making this clear, this book constitutes something of an introduction to how mysterious and mystical intuitionist mathematics is. I think that there is a significant moral to be drawn for philosophers here—it is that the temptation to view intuitionist mathematics as a set body of mathematics to which philosophers can appeal should be avoided. Brouwer's intuitionist mathematics is both weird and mystical and has its basis in what Brouwer calls, "the empty form of the common substratum of all twoities" (p. 5) but does not (and, indeed, cannot) explain. Perhaps this can be spelled out into a compelling and deep philosophical view, but it is not immediately evident that this can be done. In any case, it certainly is not done by Brouwer in this set of lectures. And, without understanding Brouwer's mathematics in the context of such a view, it seems unlikely that a philosopher who begins by understanding only classical mathematics can come to any understanding of Brouwer's intuitionist mathematics.

JOAN WEINER
University of Wisconsin-Milwaukee

REFERENCE

HEYTING, A. (*ed.*) [1975]: *L. E. J. Brouwer Collected Works*, volume I. North-Holland.

WEATHERFORD, R. [1982]: *Philosophical Foundations of Probability Theory*. London: Routledge and Kegan Paul. £15.00. Pp. xi + 282.

It is now some forty years since Ernest Nagel's well-known monograph (Nagel [1939]) was first published and in the intervening years much work has been done in the philosophy of probability. Books and articles in plenty have dealt with particular problems and theories but no general philosophical survey has succeeded it despite the inevitable signs of age. As Weatherford informs us himself, it is the aim of his book to fill this gap in the literature: 'In frustration I resolved to write the book I could not find' (p. ix). It is unfortunate then that this book could have been written thirty years ago. A handful of works are cited repeatedly—Keynes [1921], von Mises [1928], Reichenbach [1935], Carnap [1950]—and although more recent exegetical and critical works are used nothing is said of the new directions taken since these now classical volumes first dominated discussion. The reasons for ignoring Popper's propensity theory and others are sociological, not philosophical—'they are not (yet) important enough and have not been accepted and used by sufficient numbers of philosophers, mathematicians, and statisticians' (p. 6); similarly Jaynes's objective Bayesianism is rejected as being 'historically and systematically unimportant' (p. 17). (All unprefixed page numbers refer to the book under review. Emphases are in the original unless otherwise indicated.)

The book opens with an introductory chapter which raises the question, 'What is probability?', and in which are sketched the major theories of probability. Two desiderata which any theory of probability ought to satisfy are presented. Firstly, a theory of probability must contain the familiar calculus of probability or some closely similar calculus. Secondly, a theory of probability must give an interpretation of this calculus or provide a definition which enables us to use the calculus in making probability judgments. In analysing theories of probability two questions are posed: the metaphysical one of what probabilities are, and the epistemological one of how they are known. In the sequel four types of theories are examined: the Classical Theory of Probability (CTP), the A Priori (AP), the Relative Frequency (RF), and Subjectivistic (SUB) theories, a chapter being devoted to each.

The discussion of CTP is well-balanced and for the most part unexceptionable. A brief historical sketch, culled from standard sources, comes first. Following Carnap, in opposition to Nagel, Weatherford considers CTP not to be a subjectivistic theory, rational degrees of belief being the subject matter. Bernoulli's Theorem is analysed lucidly and it is emphasised that the theorem links the probability of a repetitive event to the probability of a frequency, not a frequency itself. This is only a point of logic but too often more is claimed on behalf of this theorem. The only oddity in the discussion of CTP is the suggestion that the Principle of Indifference be interpreted in terms of random guesses. This construal is a rational reconstruction rather than a reflection of the views of proponents of CTP.

AP theories are seen as an attempt to purify and correct CTP and to solve the verification problem. As Bernoulli's Theorem does not predict a frequency which must occur CTP assignments of probability are empirically unverifiable; AP theories get around this problem by making verification a logical matter (p. 70). A sketch of Keynes's theory which includes a discussion of his use of the Principle of Indifference, is followed by an introduction to Carnap's early system. The major problem with the latter lies in the specification of criteria with which to judge *c*-functions. As Weatherford observes, Carnap evaluated *c*-functions by considering how closely they are in accord with our intuitions concerning probability (p. 115). When it comes to grips with the details of this Weatherford's account lacks coherence. Support is given to the idea that we can view a collection of *c*-functions analogously to a family of geometries (p. 121), but of a *c*-function is to be chosen for its 'resemblance' to reality, whether mental or physical, co-ordinative definitions are needed and no-one has ever provided an interpretation for a *c*-function. Explication is confused with formalisation of judgments made according to some intuitive probability theory. The distinction between formal and physical geometry is relevant here, particularly in regard to the AP theorists' claims of analyticity and apriority. AP theories have to conform to our intuitions and practice concerning probability and the reality to which *c*-functions are to be compared is a mental or behavioural non-physical reality (p. 116, p. 121). Given Weatherford's views on SUB (see below) it cannot be an individual's intuitions or behaviour, rather that of an ideally rational agent. How is one to recognise a rational agent or rational behaviour if he does not know which *c*-function prescribes rational degrees or belief? In any case, a *c*-function is part of the semantic theory of some language L, hence belongs to a (formal) meta-language. To compare *c*-functions requires a meta-meta-language. It is surely doubtful whether anyone believes that *this* is what's going on when we assess probability statements. For all the fine rhetoric about the accord between our representation of reality and reality itself (p. 116, p. 128), no information is given in regard to how an individual's intuitions concerning probabilistic relations between propositions resemble the relationships (probabilistic relationships?) between events. The best that can happen is that, in the light of sequentially accumulated evidence, the numerical values which a *c*-function attributes to some repetitive event converge in the limit to the observed relative frequency. As is well known, this is in nearly all cases independent of the initial probability assignment. Furthermore Weatherford rejects RF theories of probability so even this relative frequency is not the probability which we are after. It appears that if one accepts only an AP theory and considers some probability statement to be true then it is true in virtue of something which one cannot possibly understand for this 'something' is not an AP probability.

Weatherford's tone in his account of RF theories, particularly von

Mises's, betrays a marked lack of sympathy. Conjectures about von Mises's philosophical beliefs are made which contradict his stated views. For example Weatherford claims that von Mises's contention that every measurement involves a collective and limits 'obviously turns heavily on the verificationist (or operationalist) tenet that the *meaning* of a physical statement (or quantity) depends on how it is verified (or measured) in practice' (p. 183). Von Mises explicitly rejected the verificationist theory of meaning: 'It does not appear expedient to accept verifiability (or perhaps in the unilateral sense, falsifiability) as the sole criterion of a sentence's admissibility, because the question of verification depends upon the linguistic rules' ([1939] p. 76). Moreover Weatherford undermines his own argument when he says that 'if an empiricist adhered to the strict interpretation of the Logical Positivists' Verification Theory of Meaning, he would be forced to abandon RF theory as being meaningless because not completely verifiable in experience' (p. 205). Weatherford submits that 'von Mises probably thought he was doing for probability theory what Reichenbach had done for geometry' (p. 156), and goes on to say that this 'would require von Mises to exhibit probability theory as a pure mathematical theory whose validity is independent of *any* contact with the real world', but that he 'could not bring himself to do this, since he *also* wanted to think of probability theory as an inductive science, firmly grounded on empirical fact'. However von Mises regarded probability theory and geometry as physical theories just as much as are thermodynamics and mechanics: 'The calculus of probability, . . . like geometry, mechanics, etc., is the exact theory of a specific area of phenomena' ([1939] p. 174).

As to unsympathetic interpretations, Weatherford quotes von Mises's remark that 'the exclusive purpose of the theory of probability is to determine, from the given probabilities in a number of initial collectives, the probabilities in a new collective derived from the initial ones' ([1928] p. 32). He comments that von Mises' 'greatest achievement was his definition of "probability" and his explanation of the source of initial probabilities' (p. 156), and continues, 'odd that a man's fame should rest on something he denied he was doing!' (*Cf.* a similar quotation and comment, pp. 156–7.) Von Mises's remark is only intended to contrast his own theory in which initial probabilities are determined empirically, the theory tells us how, with CTP which specifies *a priori* the numerical value of initial probabilities. It is this empirical aspect to determination which is behind his claim that no determination of probability, however exact at the time it is made, remains valid forever; this has nothing to do with objective changes in the relative frequency, contrary to Weatherford's contention (p. 177). The objection that RF probabilities are eternally constant (p. 216) does not hold water: they are necessarily temporally invariant only if time does not enter into the definition of the collective (reference class). If all RF probabilities were eternally unchanging no-one could claim that RF theories are adequate to scientific and acturial uses.

Apart from these points Weatherford's criticisms of RF theories are for the most part justified and standard. The objection, attributed to Kneale, that RF theories confuse the evidence for, and manifestation of, a probability with the probability itself is indeed forceful and is supported by von Mises's remark to the effect that 'the probability of a 6 is a physical property of a given die' (quoted p. 183). It is exactly this confusion which the propensity variant of RF aims to avoid. It should, however, be remembered that in acturial uses the relative frequency *is* the probability. To reject RF theories *in toto* because of this confusion in practice (p. 250) is to react in a philosophically uncritical manner.

The chapter on SUB theories is less than one third as long as the discussions of CTP, AP and RF theories, and is quite inadequate. Weatherford claims that 'SUB gives no rule for connecting past experience to present probabilities' (p. 230). He would appear therefore to be ignorant of the baroque creations of probability kinematics, whose aim is the evaluation of just such rules. Moreover, were his claim true it would apply equally to AP theories which would then provide no basis for inductive logic. He is intent on estranging AP and SUB whereas it is their similarity which accounts for the existence of objective Bayesianism. In the cursory examination of SUB theorists' views of physical chance the de Finetti Representation Theorem, the key to any such analysis, is not mentioned. SUB theorists are accused of psychologism. 'The most basic criticism of subjectivistic probability is that it confuses feeling with fact' (p. 238). This is simply false. SUB theorists do not *confuse* feeling with fact, they *deny* the existence of the fact, and even the meaningfulness of talking about it. And it is precisely this which Weatherford finds unacceptable. The 'killer' argument runs as follows. 'Few of us are so soft a mark that we accept certain losses. But many of us believe in the maturity of odds, betting a martingale, or drawing to an inside straight. The gambler doesn't cheat or deceive us, and we need not be incoherent fools to lose our money—the gambler understands the odds and plays the percentage and, in the end, will always take our money *because the odds are real*' (p. 238). Similarly, in considering two people betting on the throwing of a die at odds of $\frac{1}{2}$ and 1/6, we are told that 'it is perfectly plain that one will lose his shirt *because he is wrong about the probability*' (p. 247). But that one will lose his shirt can only be known (or better, seem likely) on the basis of past experience, *i.e.* on the basis of a decision as to whether the event about to occur is a repetition of a previous event. That this is a decision and not a matter of fact is true not only for the subjectivist but for everyone. AP theories present the illusion of objectivity only if one fails to remember that the descriptive adequacy of the object language has to be determined.

The final chapter of the book, *Conclusions*, is a delightful pot-pourri of half-baked ideas. CTP is rejected because it is not a theory of probability, *i.e.*, it doesn't tell us what probability is. RF theories are rejected because of the confusion of probabilities and evidence for probabilities. SUB theories are rejected because they deny the reality of probabilities. Probability theory works therefore probabilities are real (p. 250). They are real in the sense that probability statements are objectively true or false: Weatherford invokes Kreisel's distinction between epistemological realism and ontological realism, of which the former alone suffices to guarantee the objectivity of mathematics (Platonism without forms, truth without satisfaction). So let's thank our lucky stars we've still got AP! 'But Keynes is too intuitive and Carnap too abstract to capture exactly what I mean.... Probability theory is closer to logic than von Mises would have it, but *logic is closer to science than Keynes ever dreamed*' (p. 251). This closeness is accounted for by the 'Harvard view' of the *a priori*, according to which even logic is not immutable. Weatherford cites his 'paradigm case of good probabilism'—Bose-Einstein statistics (pp. 251–2). Bose and Einstein produced an *a priori* model, distributing particles in state-descriptions according to the Rule of Indifference, and tested their model using relative frequencies as evidence for the actual probabilities. They discovered that the model applied to some particles but not to all. Quantum mechanics explained the results later. Why Weatherford should consider this a particularly good example is not clear. His closing remarks inform us that 'no adequate philosophical explication of probability theory exists as yet, nor can one do so until we learn a good deal more logic and a good deal more physics than we presently know' (p. 252). Fortunately for those of us who take a more optimistic view of the philosophy of probability Weatherford provides no argument to support this disheartening conclusion.

This book may serve, in the manner of Wittgenstein's ladder, as an introduction to philosophical problems concerning probability for students of philosophy of science. It includes a useful bibliography of works in English but will be of no great interest to specialists. No mathematical sophistication is required and only knowledge of the axioms of the probability calculus presupposed.

Corrections. p. 39: Bernoulli's Theorem takes the form

$$\lim_{N \to \infty} P(|p-F| < E) = 1,$$

where F is a function of N. p. 90: $c\dagger(h,e) = \frac{1}{4}$. p. 154: *Modus ponens* is a rule of inference, not a formula. p. 203: The first members of the series specified by

$$S_N = 1 + \frac{1}{N+1}$$

are 2, 3/2, 4/3, 5/4, 6/5, 7/6, ...

PETER MILNE
London School of Economics

REFERENCES

CARNAP, R. [1950]: *Logical Foundations of Probability*. Chicago.
KEYNES, J. M. [1921]: *A Treatise on Probability*. London.

MISES, R. VON, [1928]: *Wahrscheinlichkeit, Statistik und Wahrheit.* Vienna. English translation, *Probability, Statistics and Truth,* London, 1957. Page references to this edition.

MISES, R. VON [1939]: *Kleines Lehrbuch der Positivismus.* The Hague. English translation, *Positivism: A Study in Human Understanding,* Cambridge, Mass., 1951. Page references to this edition.

NAGEL, E. [1939]: *Principles of the Theory of Probability.* Chicago.

REICHENBACH, H. [1935]: *Wahrscheinlichkeitslehre.* Leiden. Expanded English translation, *The Theory of Probability,* Berkeley, 1949.

ANNOUNCEMENT

PHILOSOPHICAL INQUIRY, the international philosophical quarterly, founded in Athens (1978) by Professor D. Z. Andriopoulos, begins a second period of publication. It is now located in New York: Department of Philosophy, Herbert H. Lehman College, City University of New York, Bronx, N.Y. 10468. Its Editorial Board is reorganized and draws upon the expertise of 60 philosophers in the Universities of 50 countries. To its Honorary Board, along with Professors A. Ayer (Oxford), Gr. Vlastos (Berkeley-Cambridge), G. de Vogel (Utrecht), W. H. Walsh (Edinburgh), G. Calogero (Rome), P. Agaesse (C.E.R.P. Paris), H. Osborne (London), R. Kromer (Switzerland), were added last Spring Professors N. Chomsky (M.I.T.) and W. V. Quine (Harvard). Professor A. McLaughlin (City University of New York-Lehman College) will serve as Associate Editor of the journal.

The Journal of Philosophy

Subscriptions at $15.00/year; $12.00 to students
$20.00/year to libraries

Students' price extended to retired and unemployed philosophers. Published bimonthly 1904 to 1976, monthly thereafter. Complete volumes and all separate issues available back to January 7, 1904 (volume I, number 1); prices as follows:

BACK VOLUMES (from the first to the penultimate year)
32-page issues, $1.50; 64-page, $3.00; special issues, $4.00.
Volumes (unbound, volume indexes included), $30.00 each.

CURRENT VOLUMES (this year and last)
64-page issues, $2.00; special issues, 96-page or over, $3.00.
Volumes (unbound, volume indexes included), $25.00 each.

ALSO AVAILABLE
Cumulative Fifty-year Index, 1904–1953; articles classified by subject and author; 452 p.; cloth, $12.00. Ten-year Supplement, 1954–1963: 98 p.; cloth $3.00, paper $2.00.

720 PHILOSOPHY HALL, COLUMBIA UNIVERSITY, NYC 10027

MIND

A QUARTERLY REVIEW
OF PHILOSOPHY
edited by D. W. Hamlyn
with the assistance of S. D. Guttenplan

Vol XCIII No 370 April 1984

Articles

Seeing the Present: J. BUTTERFIELD
Rigidity and Scope: A.D. SMITH
On the Nature of Rights. J. RAZ
Events and Explicative Definitions. G.N. SCHLESINGER
An Evidentiary Theory of Promises: J.P.W. CARTWRIGHT

Discussions

The Principle of Equality: A Reply to Nathan. T. HONDERICH
Lowe on Modalities *De Re*· T. BALDWIN
Frege on Thoughts: A Reply: G. CURRIE
Why Nozick is a Sceptic: A.L. BRUECKNER
Why Time is Extensive: G. PLUMER
Consciousness and the Exclusivity of Function: E R. KRAEMER

Critical Notice

The Interpretation of Frege's Philosophy by Michael Dummett. D.R. BELL

Book Reviews **Books Received** **Notices**

Orders to Journals Department,
Basil Blackwell,
108 Cowley Road,
Oxford OX4 1JF, England

Annual subscription
Institutions: £14 50 or $32 50 (Canada $37.75)
Individuals. £10 00 or $22 50 (Canada $26.25)
Single issues. £4 00 or $9.00 (Canada $10 50)

Published for the MIND ASSOCIATION by BASIL BLACKWELL

The Philosophical Quarterly

Vol.34 No.134 JANUARY 1984
© The Philosophical Quarterly 1984

Edited in the University of St. Andrews by an Editorial Board
Chairman of the Board of Editors: LESLIE STEVENSON
Executive Editor: GORDON GRAHAM *Reviews Editor*: ANTHONY ELLIS

Emotion and Imagination	*John Casey*
Vico's Theory of Knowledge	*Perez Zagorin*
Dummett on Temporal Operators	*Jeremy Butterfield*
Nozick on Scepticism	*Graeme Forbes*
DISCUSSIONS	
Conjectural Realism	*Richard Jennings*
Brown in Defence of Scientific Realism	*Roland Puccetti*
Reply to Puccetti	*J. R. Brown*
Mill and the Utility of Liberty	*Mark Strasser*
BOOK REVIEWS	

Subscriptions for 1984

Annual Subscription	UK	Overseas	USA	Canada
Institutions	£22.50	£27.50	$50.00	$62.00
Individuals	£10 25	£12.50	$25.00	$30.00

Orders, remittances and enquiries should be addressed to: Journals Department,
Basil Blackwell Publisher Ltd., 108 Cowley Road, Oxford OX4 1JF, England

---

Noûs

Hector-Neri Castañeda, *editor*
Earl Conee, *associate editor*

The primary purpose of *Noûs* is to publish outstanding essays that extend the frontiers of philosophical research. *Noûs* emphasizes positive theoretical work that takes full account of current developments in philosophy. There are no editorial restrictions on either the area of philosophy that is discussed or the philosophical orientation of the discussion. *Noûs* offers several distinctive services to its readers:

CRITICAL REVIEWS: Each issue contains detailed evaluations of a few important recent philosophical books.

WESTERN DIVISION SYMPOSIA PAPERS: The March issue of each volume is devoted to the invited addresses for the year's American Philosophical Association Western Division meetings.

SPECIAL TOPIC ISSUES: Frequently an issue will contain a collection of major papers on a single central philosophical topic, often guest-edited by a leading authority on the topic.

Subscription Rates for Volume XVIII (1984) only: Individuals: 1 year $16.00, 2 years $31.00, 3 years $46.00. Institutions: 1 year $32.00, 2 years $64.00, 3 years $96.00. Please add $3.00 postage per year for all foreign subscriptions.

INQUIRIES SHOULD BE SENT TO: *Noûs*, Department of Philosophy, Indiana University Sycamore 126, Bloomington, Indiana 47405 USA.

RATIO

edited by
MARTIN HOLLIS

Volume XXV (1983) **December**

ROGER SCRUTON	Understanding Music
HEINZ-DIETER HECKMANN	What a *Res Cogitans* Is — and Why I am one
LAURENCE GOLDSTEIN	Wittgenstein and the Logico-Semantical Paradoxes
GEOFFREY BROWN	Language, Rules and Authority
E. M. ZEMACH	The Possibility of a Materialistic Language
ROBERT BERNASCONI	Review of Rudiger Bubner, *Modern German Philosophy,*
DAVID WOOD	Review of Vincent Descombes, *Modern French Philosophy*
NICHOLAS EVERITT	Review of Roderick Chisholm, *The Foundation of Knowing*

Subscriptions

Institutions	£22 95 (UK)	£28 20 (overseas)	$59 95 (USA)	$74 00 (Canada)
Individuals.	£15 50 (UK)	£19.00 (overseas)	$42 00 (USA)	$52 00 (Canada)

Orders to Journals Department, Basil Blackwell Publisher Ltd , 108 Cowley Road, Oxford OX4 1JF, England

The German edition of *Ratio* is published by FELIX MEINER VERLAG, Richardstrasse 47, 2000 Hamburg 76, West Germany

Published by BASIL BLACKWELL, Oxford

THE PHILOSOPHER'S INDEX

An International Quarterly Index To Philosophical Periodicals

Expanded Coverage

The Philosopher's Index significantly expanded coverage in 1980. In the past, current issues of *The Philosopher's Index* indexed journal articles only. However, Volume XIV (1980) increased coverage to include original scholarly books in philosophy, plus bibliographies, biographies, translations, textbooks, dissertations, and anthologies. Contributions to anthologies are indexed as well as the work as a whole.

Special Editions

The Philosopher's Index: A Retrospective Index to U.S. Publications from 1940 and *The Philosopher's Index: A Retrospective Index to non-U.S. English Language Publications from 1940* greatly enhance coverage of philosophy journals and books on a retrospective basis.

Subscription Information

The Quarterly, Paperback, $69 (Individuals: $34.50) / Annual Cumulative Edition, Hardbound, $79 (Individuals: $39.50) / *Retrospective Index to U.S. Publications*, Hardbound in 3 volumes, $195 (Individuals: $130) / *Retrospective Index to non-U.S. English Language Publications*, Hardbound in 3 volumes, $180 (Individuals: $120)

U.S. and Canada:

THE PHILOSOPHER'S INDEX
Bowling Green State University
Bowling Green, Ohio 43403 USA

Outside the U.S. and Canada:

D. Reidel Publishing Company/Dordrecht, Holland

EDITORIAL BOARD

R. B. Braithwaite
G. Buchdahl
L. J. Cohen
Sir John Eccles
M. B. Hesse
Sir Harold Jeffreys
C. W. Kilmister
W. C. Kneale
S. Körner
D. J. O'Connor
Sir Karl Popper
H. R. Post
J. W. N. Watkins
G. J. Whitrow

EDITORIAL ADVISORY PANEL

The Editor is helped by members of the Editorial Board, members of the Committee of the British Society for the Philosophy of Science and by an Advisory Panel which includes:

Peter Clark
Jon Dorling
Arthur Fine
Allan Franklin
Ian Hacking
Colin Howson
Noretta Koertge
Hugh Mellor
David Miller
Alan Musgrave
John Worrall
Elie Zahar

THE BRITISH SOCIETY FOR THE PHILOSOPHY OF SCIENCE

Committee 1983-84

President: Dr. J. Giedymin
Hon. Secretary: Mr G. M. K. Hunt
Hon. Treasurer: Dr. K. V. Wilkes
Members: Dr. W. H. Newton-Smith, Dr. D. Papineau, Professor H. R. Post, Dr. P. M. Williams

The British Journal for the Philosophy of Science is the official organ of the Society, whose purpose is described by the following statement from its Constitution and Rules:

'The purpose of the Society is to study the logic, the method and the philosophy of science as well as those of the various special sciences, including the social sciences.'

The Society holds meetings, at which papers are read and discussed, once a month during the academic year. These are held, by kind permission, in University College, London. The Committee may recognise Regional Groups of Members formed to hold meetings in other parts of the country. Any Regional Group so recognised may nominate a representative Member to the Committee.

Ordinary membership is open to all interested persons on payment of the annual subscription. Membership of the Society is international and is by no means confined to people resident in the United Kingdom. The annual subscription is £4.00 for U.K. residents and £2.00 for overseas residents. Those who wish to join should write to the Hon. Secretary, Mr G. M. K. Hunt, Department of Philosophy, University of Warwick, Coventry CV4 7AL.

Brit. J. Phil. Sci. Vol. 35, No. 1, March 1984

The British Journal for the Philosophy of Science

18 JUL 1985

Contents

Articles

1–10 PETER GÄRDENFORS The Dynamics of Belief as a Basis for Logic

11–23 HUW PRICE Mellor, Chance and the Single Case

25–45 ABNER SHIMONY Contextual Hidden Variables Theories and Bell's Inequalities

47–53 ROY A. SORENSEN Problems with Electoral Evaluations of Expert Opinions

Discussions

55–57 DOUGLAS EHRING Probabilistic Causality and Preemption

57–60 ALLAN FRANKLIN Are Paradigms Incommensurable?

60–61 IAN MCCAUSLAND Einstein and Special Relativity: Who wrote the Added Footnotes?

62–64 CHARLES NISSIM-SABAT A Gedankenexperiment to Measure the One Way Velocity of Light

Review Article

65–80 JOHN WORRALL An Unreal Image (review of BAS C. VAN FRAASSEN: *The Scientific Image*)

Reviews

81–100 Vernon Pratt on V. J. JENSEN AND R. HARRÉ (*eds*): *The Philosophy of Evolution*; Leslie Stevenson on R. NOZICK: *Philosophical Explanations*; James M. Brown on ANTHONY O'HEAR: *Karl Popper*; Joan Weiner on D. VAN DALEN (*ed.*): *Brouwer's Cambridge Lectures on Intuitionism*; Peter Milne on R. WEATHERFORD: *Philosophical Foundations of Probability Theory*.

Announcement

Printed in Great Britain at Aberdeen University Press

Volume 35 Number 2 June 1984
BJPIA 35 (No. 2) 101–200 (1984) ISSN 0007–0882

The British Journal for the Philosophy of Science

Published for the
British Society for the Philosophy of Science
by Aberdeen University Press

THE BRITISH JOURNAL FOR THE PHILOSOPHY OF SCIENCE

Editor: DR. DONALD A. GILLIES

Subscriptions

The *Journal* is published quarterly. The subscription price for mailing addresses in Britain is £14.00 net per volume post free and for mailing addresses outside Britain £17.00 (or foreign currency equivalent calculated at the day's rate); single issues £5.00 net (overseas £6.00) plus postage. Orders may be sent to any bookseller or direct to Aberdeen University Press, Farmers Hall, Aberdeen, Scotland AB9 2XT.

For members of the British Society for the Philosophy of Science—who should order direct from Aberdeen University Press—the subscription price is £9.00 (overseas £12.00). Membership of the Society is international and is by no means confined to people resident in the United Kingdom.

Microform

The current volume of the *Journal* is available in microform to subscribers to the printed version, *at a combined print/microfiche rate of* £19.50 (*overseas* £22.50).

Back Volumes

Volumes 1–16 may be obtained from Kraus Reprint Ltd., FL-9491 Nendeln, Liechtenstein or Kraus Reprint Corporation, 16 East 46 Street, New York, N.Y. 10017. Volumes 17–34 may be obtained from Aberdeen University Press, Farmers Hall, Aberdeen, Scotland AB9 2XT. Prices on application. Volumes 1–33 are also available in microform from Kraus Reprint Corporation.

Typescripts and Review Books

Typescripts, correspondence and books for review should be sent to the Editor at the Department of History and Philosophy of Science, Chelsea College, University of London, Manresa Road, London SW3 6LX.

Review books are accepted on the understanding that, although they will appear in the 'Recent Publications' list, a full review cannot be guaranteed, and that books not reviewed are not returnable.

Notes to Contributors

Three copies of all contributions must be sent. These need not, on initial submission, be in any particular style, provided they are typewritten and clearly legible. However, contributions accepted for publication may have to be wholly or partly retyped to accord with the *Journal*'s usual format—details of which may be obtained from the Editor. Contributions are considered on the assumption that they have not been published, wholly or in part, elsewhere.

Articles should be accompanied by abstracts of about 100 words.

Material is accepted for publication in the *Journal* on the condition that copyright is assigned to the British Society for the Philosophy of Science.

© British Society for the Philosophy of Science 1984

For permission to reproduce material from *The British Journal for the Philosophy of Science,* please apply to the Editor.

Brit. J. Phil. Sci. 35 (1984), 101–113 *Printed in Great Britain*

Relationism and Possible Worlds[1]

by J. BUTTERFIELD

Relationism claims that our physical theory does not commit us to spacetime points. I consider how a relationist might rewrite physical theories without referring to spacetime points, by appealing to possible objects and possible configurations of objects. I argue that a number of difficulties confront this project. I also argue that a relationist need not be Machian in the sense of claiming that objects' spatiotemporal relations determine whether any object is accelerating.

1 *Introduction*
2 *Replacing Points with Material Objects*
3 *Replacing Points with Events*
4 *Dynamical Spacetimes, Possible Points and Strong Relationism*
5 *Machianism*

1 INTRODUCTION

By 'relationism' I mean the doctrine that our physical theory does not commit us to the existence of spacetime points, or spatial points and temporal instants. I call the opposite doctrine 'substantivalism' about spacetime, space or time. For brevity, I shall refer only to spacetime points, which I shall call points; what I have to say will also hold for spatial points and temporal instants.

At first sight, relationism is an attractive doctrine: points, being unobservable and immaterial, are curious objects to believe in. Yet discussions of relationism mostly grant substantivalism an advantage in that as usually presented our physical theories quantify over points (Earman [1970a]; Sklar [1974], pp. 164–5, 206; and implicitly, Earman [1970]; Weingard [1975]). Most think, and I agree, that ineliminable quantification over a sort of object entails ontological commitment to it (*cf.* Quine [1963], ch. 1). So relationists must show how to rewrite physical theories without quantifying over points, though they may of course refer to anything they consider sufficiently material to be accepted as a *bona fide* object. Doing without such quantification while capturing the content of our theories would be no mean feat: for example, it is not clear how one could attribute a geometry to a vacuum without saying that certain points sustain certain geometric relations to one another.[2]

---

[1] This is a shortened version of a paper read at seminars in Chicago, London and Cambridge. I would like to thank those present, and especially Michael Gardner, David Malament, Hugh Mellor, Michael Redhead, and Timothy Smiley, for comments.

[2] The claim that the relationist must rewrite our theories can be challenged. van Fraassen [1980] holds that the only *belief* involved in accepting a theory that quantifies over points is the

Received 9 September 1982

This paper has two aims. The first, which occupies Sections 2 to 4, is to use possible worlds to look at one way the relationist might try to rewrite our theories, *viz.* by appealing to possible (acceptably material) objects or possible configurations of such objects. The second aim is taken up only in Section 5; there I argue that relationists are not committed to Machianism.

Some points about the first aim. The idea that the relationist should appeal to possibility is a familiar one. For example, it is often suggested that the relationist should analyse talk about vacua in terms of counterfactuals about the relations objects (whether actual or possible) would sustain if they were in the relevant region (Hooker [1971], Sec. 6; Sklar [1974], p. 170; Newton-Smith [1980], pp. 39 ff.). Similarly, the relationist must presumably analyse the extendedness of material objects in terms of the spatial relations of these objects' parts; and he may need possible partitions of the objects as well as actual ones (Hooker [1971], Sec. 3). Yet the details of a relationist theory are usually avoided; sometimes it is assumed that they would be straightforward to fill in (Bennett [1974], pp. 145–8; Horwich [1978], Sec. 6). I shall argue however that if the relationist tries to fill in these details by using possible worlds, he faces difficulties. And to that extent substantivalism is supported.

The relationist might then ask why I have saddled him with possible worlds. After all their claim to explicate the notion of possibility is strongly disputed. I believe they do explicate it, so that the relationist should use them in appealing to possibility. But as I cannot hope to show that here, I will only say that possible worlds do not commit us to some contentious doctrines (such as possibilia or essences) which have caused suspicion of them. More specifically, possible worlds do not commit us to possibilia; nor to general essences in any contentious sense, such as different objects having different necessary properties; nor to individual essences, *i.e.* objects having necessary properties such that no object could have the property without being that object; nor to all possible worlds being ontologically 'on a par' with the actual one. Of course, whether to admit possibilia or essences etc. must eventually be decided. I only claim that possible worlds are flexible enough not to force one's hand over these issues. This I think is clear from the literature.[1]. So relationists who are uneasy about possibilia or essences etc. can still use possible worlds.

---

belief that the theory accounts for the observable phenomena. On this view substantivalism is given no advantage by the usual formulations of our theories; rather, relationism follows from a general rejection of theoretical entities. I do not agree with van Fraassen, but I cannot take him on here.

[1] Thus Hughes and Cresswell ([1972], p. 146) and Lewis ([1968]; [1974], pp. 39–41) respectively reject and admit possibilia. Parsons ([1967]; [1969]) argues for non-commitment over general essences. Kripke ([1972], pp. 264–72) and Kaplan ([1975], Section 4) argue that possible worlds do not commit one to individual essences. And Kripke ([1972], pp. 264 ff.), Adams ([1974]), Stalnaker ([1976]) and Hellman and Thompson ([1977], Section 6) urge against Lewis ([1974], ch. 4.1) that possible worlds do not commit us to worlds' ontological parity.

2 REPLACING POINTS WITH MATERIAL OBJECTS

The relationist project of rewriting physical theories without points is a tough one. To get it off the ground I will assume that the relationist, instead of wholly dispensing with points, constructs them out of actual and possible objects. The reason for this is familiar. Often we can only rewrite a theory while avoiding some of its objects by identifying these objects with constructions from our favoured objects; compare the reduction of arithmetic to set-theory by identifying numbers with certain sets.

Admittedly, a variety of relationism which does not involve constructing points has been suggested. Earman ([1977], pp. 102–4) and Geroch ([1972]) suggest that a relationist should regard a certain abstract algebra as the fundamental reality—points and the structures associated with them merely providing representations of this algebra. But personally I cannot believe that the world is really an abstract algebra; and Earman and Geroch admit that there are interpretative problems here (Geroch [1972], p. 274; Earman [1977], p. 110). So I think relationism's best chance lies in the less radical project of constructing points from other objects. I will however share with Earman and Geroch the assumption that the relationist is not a nominalist; I will assume he accepts such abstract objects as sets and numbers.[1]

Given these two assumptions, the relationist has some hope. For many different theories, such as Newtonian dynamics and gravitation theory and special and general relativity,[2] can be presented in a similar way which suggests that a single relationist strategy will do for them all.

All these theories posit a set of points, endowed with some further structure. This is always a manifold structure, together with some so-called geometric objects such as a connection and scalar vector and tensor fields. The geometric objects represent the metric structure of space and time and fields like the electromagnetic field. Essentially they are functions associating sequences of real numbers with points. The connection and manifold structures together allows us to differentiate these functions. Apart from positing this structure, the theory only requires the geometric objects to obey some differential equations called field equations. So we can specify such a theory by the set of n-tuples $\langle M, F_1, \ldots, F_m \rangle$, where M is a manifold of points and $F_1, \ldots, F_m$ are geometric objects on M of the sort described by the theory; and where the F_i obey the theory's field equations. Each n-tuple represents a physically possible world according to the theory; I will call these n-tuples the *worlds* of the theory.[3]

---

[1] He will not need to accept these if Field ([1980]) is right in thinking one can build tensors etc. without such objects; but for difficulties with Field's claim, *cf.* Malament [1982*a*].

[2] Here the Newtonian theories can be taken with or without preferred rest; and if without it, with or without geometrised gravity. Relativity can be taken to include electrodynamics. For these theories, *cf.* Earman and Friedman ([1973], pp. 331 ff.); Friedman ([1983]); Misner *et al.* ([1973], chs. 2, 3, 12, 17 and 22). I will consider only such non-quantal theories.

[3] *Cf.* van Fraassen ([1980], ch. 3) for a discussion of presenting theories by means of their worlds (usually called models). Though this presentation is convenient, nothing in my argument turns on it; I could take theories in the more traditional way as deductively closed sets of sentences.

If we are given the set of points, the manifold structure and the various sorts of geometric object of any such world can be defined using set theory and the real numbers. (These definitions can be reconstructed from differential geometry texts, *e.g.* Bishop and Goldberg [1968].) So since the relationist accepts sets and numbers, he can build worlds corresponding to those of the substantivalist theory, provided he finds some replacement for the points. Accordingly from now on I will only consider how the relationist replaces points.

Since the relationist wants to reduce points to material objects and their relations, the first question is: what counts as a material object? When Leibniz and Clarke had their famous debate (Alexander [1956]), the distinction between empty space and material objects seemed much firmer than it does today. In particular, the distinction could then have been made in terms of the absence or presence of mass. And other attributes went along with this distinction: only objects with mass could have energy and momentum. But since then classical physics has introduced the electromagnetic field endowed with energy and momentum; and relativity has identified mass and energy. Furthermore, in general relativity there can be material in the sense of mass-energy in a region where there is not only no material object of an ordinary sort like a chair, but also no field apart from the metric-gravitational field.[1]

This widening of the notion of material apparently vindicates the relationist's traditional rejection of vacua. And in that case, one might say, the relationist need not appeal to possibility: points can be eliminated in favour of the extensionless material objects which—as the substantivalist would say—occupy them. I shall call such objects *events*.

However, vacua cannot be got rid of so easily, for two reasons. Firstly, not every world of general relativity (or any other of the theories listed on p. 103) is everywhere filled with material, in even the weakest sense of 'material'.[2] So a relationist trying to rewrite these theories must still deal with vacua. I admit that this point might be countered by suggesting a principle for rejecting worlds that are not everywhere filled with material. After all, we should not require the relationist to produce a theory which perfectly matches the substantivalist's; he should be allowed to reject some of the substantivalist's worlds as unreconstructable from his point of view. But this reply will not work against the second reason why vacua cannot be got rid of. This is that a strong-minded relationist will find the weak notions of material that promise to eliminate vacua unacceptably weak. He will advocate some stronger notion of material as the arbiter of what is real, and thus admit that vacua occur in some substantivalist worlds that he has to reconstruct.

---

[1] Admittedly, gravitational mass-energy cannot be localised (Misner, Thorne and Wheeler [1973], pp. 466–8). But it is real enough, as any gravitational wave detector shows; *e.g.* one can heat water with it (*ibid.*, pp. 444–5, 1004 ff.).

[2] Since gravitational mass-energy cannot be localised, the identification of points with events will be most plausible if we take this weakest sense to be the non-vanishingness of the stress-energy tensor T_{ab} rather than the presence of gravitational mass-energy. This point only makes the objection stronger, since T_{ab} can be zero in regions with gravitational mass-energy.

I will take it then that the relationist must deal with vacua, with unoccupied points; and that to do this he will appeal to possibility. Can he at least maintain the suggested identification between occupied points and the events which—as the substantivalist would say—occupy them?

One can present two objections to this. Firstly, if events are to be identified with the points they occupy, the same event must be prevented from occurring at distinct points within a world; and it is not easy to say why they cannot do so. The relationist cannot use events' spacetime locations to identify them. But if he doesn't do that, he must face the fact that the most natural notion of event-identity—sameness of value for some or all fields—*does* allow the same event to occur at distinct points within a world.[1] Secondly, a strong-minded relationist will want his theory to refer only to extended objects, and so will find these extensionless events unacceptable.

However I will ignore these objections. I will suppose that the relationist accepts events, and can produce a notion of event-identity which prevents the same event occurring at distinct points within a world. This simplification will clear the way for considering other difficulties facing the relationist. And in any case the relationist has good reason to accept extensionless objects like events. The commitment to rewrite physical theory using only extended objects is very ambitious. Even in elementary geometry one cannot say much while referring only to extended objects (Hooker [1971], Sections 3, 4, 6; Russell [1901]).[2]

3 REPLACING POINTS WITH EVENTS

At this stage the natural relationist move is to take the occupied points in each world of the substantivalist theory to be events; and to take the unoccupied points of each world to be possible events; that is, events which do not exist in that world but which, as the substantivalist would say, could occur at these points—*i.e.* do occur at them in other substantivalist worlds.

If this is the relationist's position, the dispute between him and the substantivalist will in effect be about whether 'exists', 'actual' etc. should apply to more than just the occupied points. For in effect the substantivalist is taking as his points, all existing in the actual world, the union over all possible worlds of the relationist's events. A similar question arises more generally about domains of quantification in modal logic. Once we admit possibilia, the question arises whether they should fall in the actual world's domain of quantification, actual objects being distinguished from them by an actuality or existence predicate; or whether each world's domain should contain only the objects that are actual in it. The analogy here is plain: the substantivalist is advocating the first alternative, the relationist the second. So we might hope to settle the substantivalist-relationist dispute by reducing it to the more general one about existence.[1]

---

[1] 'Sameness of value' here will no doubt be interpreted using parallel transport along a favoured geodesic; the notion can certainly be applied locally (Hawking and Ellis [1973], p. 34).

[2] One can of course construct extensionless objects mereologically in various ways, *e.g.* in terms of sequences of extended objects whose 'widths' tend to zero (Whitehead [1920], pp. 84–93; Tarski [1956], pp. 24–9). But I think the cure is as bad as the disease; that is, positing a set of extended objects with the required structure seems as questionable as positing extensionless objects *simpliciter*.

However, we cannot in fact make such a reduction. The proposed identification of points with events faces at least three difficulties. The relationist might at first be heartened by these difficulties; as he may want his theory to be more different from the substantivalist's than the above reduction implies. But his optimism will not last long; for each of these difficulties makes it less clear how the relationist could rewrite the substantivalist's theories. I will present one difficulty in this Section, leaving the others to Section 4.

The first difficulty is that the proposed identification of points with the events at them may not hold across all worlds. The substantivalist may well allow some points to have different events at them; that is, there will be at least one point which is occupied in two worlds by different events. Taking the union of the relationist's domains for these two worlds, we get two events for just one point, disturbing the proposed identification. Conversely, we may want the same event to occur at two distinct points in two worlds. (I supposed in Section 2 that an event could not occur at two points within a world; but this allows events to occur at two points in two distinct worlds.) In short: the relation 'occupies in some world' between events and points is many-many.

The obvious solution to the first of these two problems—one point having different events—is to identify points with the set of events that occupy them in some world or other. That is, points are taken as equivalence classes of the union over all worlds of the relationist's events; the equivalence relation being 'is according to the substantivalist at the same point as'. But this solution means of course that an event cannot occur at two points in distinct worlds. If it could, distinct points might have the same associated equivalence classes and thus have to be identified.[2]

---

[1] This description of the two treatments of possibilia is not precise; for example, it leaves open how to treat negation. But we may nevertheless feel that the dispute between these two sorts of treatment is verbal, *i.e.* that the notion of existence is not precise enough to establish a preference between them. In that case the substantivalist-relationist dispute, if reduced to this one, would also be verbal—as has been suspected (Malament [1976], p. 317).

[2] In particular, we can hardly rely on our physics to yield the result that, as the substantivalist would put it, for any two points there is an actual or possible event that could occur at the one but not at the other.

4 DYNAMICAL SPACETIMES, POSSIBLE POINTS AND STRONG RELATIONISM

At least two other difficulties face the proposal that occupied points are actual events and unoccupied points are possible events. They also face the amended proposal that points be identified with equivalence classes of actual and possible events. For brevity I will present them as difficulties for the amended proposal; it will be obvious that they also face the original proposal.

One difficulty is that it may be impossible to occupy some points. Clearly such a point could not be identified with the set of events that could occupy it: at least, this is so if there are two such points, for then each would have to be identified with the null set and thus with the other.

The relationist will naturally question the claim that some points cannot be occupied. But this claim is plausible when we consider a theory like general relativity, or Newtonian mechanics with geometrized gravity, which is dynamical in the sense that the metric and/or connection is affected by matter. This means that the metrical structure near an event will generally differ from what it would have been had the event not occurred. That is, two worlds of the theory which differ as little as is physically compatible with the event occurring in one but not the other, will differ in the metrical structure of the event's neighbourhood. Now, the substantivalist may well take metrical isomorphism—a bijection preserving metrical structure—of some two neighbourhoods of two points to be a necessary condition of the points' identity. And in that case, there may well be points that cannot be occupied; for their being occupied would destroy the metrical isomorphism that is necessary to occupying *them*.

Another difficulty arises if the substantivalist admits possible as well as actual points. And this he is likely to do: for example, he will have to do it if a point in one world has properties so different from those of the points in another world that it cannot be identified with any of them.[1] If the substantivalist has possible points, the relationist will have difficulty in expressing the distinction between actual and possible points. It is now of course the actual and possible points which are identified with equivalence classes of the union over all worlds of events. But we cannot pick out the actual points from the possible ones as the equivalence classes that contain an actual event. These equivalence classes correspond rather to the actual *and occupied* points; for we are assuming that each event can occur at only one point, so that the actual events can occur only at the points they do occur at—the actual and occupied points. Nor can we pick out the actual points from the possible ones as the equivalence classes whose elements could occur in the actual world; since all possible events could occur, that would make all points actual, which *ex hypothesi* they are not. Clearly, to pick out the actual points we must distinguish somehow between what could occur, and what could occur while retaining the identity of the actual points. But it is not clear how the relationist can do this.

---

[1] It is indeed true that every world of every familiar substantivalist theory has continuously many points (Geroch [1968], p. 1743) so that the actual points will be numerous enough to serve as the points of any world. But this does not avoid possible points, unless we add a contentious assumption, *e.g.*: the actual points are physically necessary existents, and their only essential property is being a point. With this assumption, any world can use the actual points as the locations of whatever events it contains.

This completes my survey of the difficulties that inevitably confront the relationist who appeals to possibility in order to avoid points. I now present a difficulty that confronts only a certain variety of relationism, albeit a traditionally popular one: I call it strong relationism.

Strong relationism adds to the basic relationist rejection of points the claim that certain possibilities which the substantivalist regards as distinct are in fact identical. The strong relationist identifies worlds which do not differ in the properties and relations of material objects, even if according to the substantivalist they differ in the location of these objects in space and time. Thus traditionally he identifies worlds which differ by a translation of all objects in time and/or space, and/or a rotation of all objects in space, and/or a constant velocity boost of all objects. (Compare Leibniz's appeal to the Principle of Sufficient Reason and to the Identity of Indiscernibles (Alexander [1956], pp. 26–7, 38–9).)

If the strong relationist appeals to possibility to avoid points, he will face the three difficulties discussed above: the many-many association of events and points, dynamical spacetimes and possible points. But he will also have another difficulty. This arises from the fact that none of our previous difficulties prompted us to associate one possible relationist world with more than one substantivalist world. On the contrary, the amended proposal of Section 3 followed the general strategy of the relationist's appeal to possibility: *viz.* associating several relationist worlds with one substantivalist one. But strong relationism reverses this: it associates one relationist world with many substantivalist ones. Consequently it is even harder to see how to write down the relationist theory; for we cannot expect to recover the substantivalist theory by anything like forming quotient sets of events, and unions of domains.

5 MACHIANISM

I turn to my second aim: showing that relationism, even strong relationism, does not entail Machianism. From relationism's point of view, this is a more positive aim than the first, since it means that relationists can hope to rewrite the theories we now accept and have accepted in the past, *viz.* Newtonian dynamics and gravitation theory and special and general relativity. For none of these theories are Machian.

And what is Machianism? I shall take it to claim that the spatiotemporal properties and relations of material objects determine whether any object is accelerating. This construal fits nicely Mach's rejection of Newton's thought-experiments with the bucket and the two globes (Newton [1934], pp. 10–12; Mach [1942], pp. 279–84). Each of these experiments describes two situations in which the spatiotemporal relations of all material objects are the same, but which differ in that in one situation an object is accelerating while in the other it is not. To accept that both situations are possible would mean denying that the spatiotemporal relations of all objects determine whether any object is accelerating. And in this sense of Machianism, the theories mentioned above are non-Machian: Newton's bucket experiment is as valid in relativity as in Newtonian theory (*cf.* Earman [1970], pp. 300–2; Sklar [1974], pp. 198–202, 206–21; Gardner [1977], pp. 228–9 for more discussion).[1]

It is tempting to think that strong relationism entails Machianism. The strong relationist claims that the world is determined by material objects' properties and relations—wherever, as the substantivalist would say, these objects are in space and time. The Machian's claim about the determination of acceleration is apparently weaker, since there is more to the world than facts about accleration. I shall argue however that the entailment fails. The reason lies in the distinction between material objects' *spatiotemporal* properties and relations and their *other* properties and relations. That is, it may be that material objects' properties and relations determine whether an object is accelerating as required by strong relationism, even though their spatiotemporal properties and relations do not, thus defeating Machianism.

If we had a strong relationist theory, we could show the non-entailment by exhibiting two worlds of the theory which refute Machianism by having the same spatiotemporal properties and relations between material objects and different facts about acceleration. (This is what happens in Newton's bucket and two globes experiments.) However since we do not have a strong relationist theory, I must proceed indirectly.

I will first state a more precise version of strong relationism, and argue that this version can indeed identify worlds which strong relationists have traditionally identified, *e.g.* worlds in which material objects have the same properties and relations but which differ by a translation of all objects in space. But there will be no reason for the strong relationist to identify worlds in which material objects have the same spatiotemporal properties and relations but which differ by an acceleration such as a rotary motion. For he will have no reason to reject the conclusion of Newton's two experiments: *viz*. that such an acceleration will produce differences in material objects' *non*-spatiotemporal properties and relations, *e.g*. tension in the cord between the globes. And since the Machian must identify these worlds, this means the strong relationist need not be Machian.

---

[1] As various authors have stressed (Stein [1967], [1977]; Earman [1970]; Malament [1982]) a connection is precisely what is needed to make sense of acceleration; so one could take Machianism as the doctrine that the spatiotemporal relations of material objects determine the connection. This construal has the advantage of allowing one to argue that Newtonian theory is non-Machian without referring to Newton's bucket; one need only point out that the connection is not determined by the spatial and temporal metrics (Stein [1967], p. 187). However I can make do with the first construal. (It entails the second if it is understood to claim that objects' spatiotemporal relations determine whether a particle following any spacetime path would be accelerating; for a connection is fixed by its set of non-accelerating paths.)

To make strong relationism more precise, consider some expanded substantivalist theory which explicitly allows for material objects and their properties and relations. Its worlds will be n-tuples $\langle M,O,F_1, \ldots, F_m,v(A_1), \ldots, v(A_l)\rangle$, where M is a manifold; O is the set of material objects coincident with various points of M; the F_i are geometric objects on M satisfying the theory's field equations, and the $v(A_j)$ are the extensions of the predicates A_j defined on O, and satisfying whatever laws govern the predicates A_j. We define two such expanded worlds to be *equivalent* if they have the same set of objects O and the same extensions for each predicate A_j. Then strong relationism claims that two such equivalent worlds represent the very same situation, even though the extensions of the A_j may not determine where the objects in O are within the manifold. And thus the strong relationist theory—yet to be formulated—will have only one world for each class of equivalent substantivalist worlds.

Which worlds are equivalent thus depends on the choice of the set of predicates A_j. It will have to be a rich set: unless the A_j represent every property and relation of material objects the strong relationist considers real, an isomorphism in the A_j together with the identity of the objects O will hardly suffice to identify two equivalent worlds. But which worlds should the strong relationist consider equivalent?

We can get some insight into this by considering maps on manifolds. Any sufficiently well-behaved map f from a manifold M to a manifold M' can be thought of as placing the objects O, located in M, in M'; the objects are moved to the f-images of the points they coincide with. (M' might be M itself, in which case the objects in O are moved around in M.) One can define the extensions of the A_j on O in M' to be what they were in M. And in a similar way to f's carrying these extensions into M', one can think of it as carrying the geometric objects F_i into M', the images being written as $f_*(F_i)$ (Hawking and Ellis [1973], pp. 22–4).

The new n-tuple $\langle M',O,f_*(F_i),v(A_j)\rangle$ will be a physically possible world if (i) the $f_*(F_i)$ satisfy the field equations; and (ii) the $v(A_j)$ satisfy the laws governing the A_j. These two conditions are often expressed as: the field equations and the A_j's laws are invariant under f. And if the n-tuple is physically possible, it is of course equivalent to the first world $\langle M,O,F_i,v(A_j)\rangle$: by definition, the two worlds have the same set of objects O and the same extensions for the A_j. So we can find equivalent worlds if we can find maps f satisfying (i) and (ii).

In some cases the set of maps satisfying (i) is well-known. For the simplest classical spacetime, which is globally like $\mathbf{R}^4$ and has a flat connection, the set of maps in the Galilean group. For the simplest relativistic spacetime, which is globally like $\mathbf{R}^4$ and has a flat connection, it is the Lorentz group.[1] The maps composing both these groups correspond to a translation in time and/or space, and/or a rotation in space, and/or a constant velocity boost. The strong relationist has of course traditionally identified worlds differing by just such maps. So to endorse these traditional identifications, he has only to claim that condition (ii) is satisfied: that the A_j's laws are invariant under these maps. This he is very likely to do: only if the A_j represent such properties as having absolute position so-and-so, or having absolute velocity such-and-such, will their laws not be invariant under these maps.[2]

However, there is no reason to think that the A_j's laws should be invariant under other maps than these; so the strong relationist need not make further identifications. In particular, there is no reason to think that two worlds in which objects have the same spatiotemporal properties and relations, but which differ by a uniform or non-uniform rotary motion, should be equivalent. Admittedly, the A_j representing objects' spatiotemporal properties and relations will have identical extensions in two such worlds. But objects have other properties and relations besides spatiotemporal ones; and Neaton's thought-experiments suggest that the extensions of predicates for some other such properties and relations *will* vary between the worlds, preventing equivalence: *viz.* predicates representing concavity of the water-surface, or tension in the cord. (And of course the A_j must be rich enough to represent these properties if identity of their extensions is to be a convincing sufficient condition for the identity of worlds.)

To conclude: even a strong relationist need not be a Machian. So he need not be disheartened by the failure of Machianism in the physical theories that he has to rewrite without referring to points. What he *should* be troubled by are the problems, presented in Sections 2 to 4, which confront that project of avoiding points.

University of Cambridge

---

[1] *Cf.* Friedman ([1974]). Both groups are usually introduced in terms of coordinate transformations which fix the components of the geometric objects, rather than in terms of point transformations as here. The treatments are connected by the usual association of point and coordinate transformations: with a given point transformation $f{:}p \rightarrow f(p)$, we associate a coordinate transformation $x_i \rightarrow x_i'$ such that $x_i(p) = x_i'(f(p))$.

[2] More generally, if the F_i determine the A_j then the A_j's laws will be invariant under the isomorphisms of the F_i. 'Determination' here means as usual 'no A_j-difference without an F_i-difference'. To be precise, the F_i determine the A_j iff: for any two physically possible worlds with the same set O, $\langle M, O, F_i, v(A_j)\rangle$, $\langle M', O, F_i', v'(A_j)\rangle$, if the map f from M to M' is an isomorphism of the $F_i(f_*(F_i) = F_i'$ and similarly for the inverse) and f carries the elements of O into themselves, then $v(A_j) = v'(A_j)$, *i.e.* f carries the extensions $v(A_j)$ into themselves. *Cf.* Hellman and Thompson ([1975], esp. p. 559); like them I think it plausible that physical predicates determine in this sort of way all other predicates.

REFERENCES

ADAMS, R. [1974]: 'Theories of Actuality', *Nous*, **8**, pp. 211–31.

ALEXANDER, H. (*ed.*) [1956]: *The Leibniz-Clarke Correspondence*. Manchester: Manchester Univ. Press.

BENNETT, J. [1974]: *Kant's Dialectic*. Cambridge: Cambridge Univ. Press.

BISHOP, R. and GOLDBERG, S. [1968]: *Tensor Analysis on Manifolds*. New York: Macmillan.

EARMAN, J. [1970]: 'Who's Afraid of Absolute Space', *Australasian Journal of Philosophy*, **48**, pp. 287–319.

EARMAN, J. [1970*a*]: 'Spacetime, or . . .', *Journal of Philosophy*, **67**, pp. 259–77.

EARMAN, J. [1977]: 'Leibnizean Spacetimes and Leibnizean Algebras', in *Historical and Philosophical Dimensions of Logic, Methodology and Philosophy of Science* (*eds*. R. E. Butts and J. Hintikka). Dordrecht: Reidel.

EARMAN, J. and FRIEDMAN, M. [1973]: 'The Meaning and Status of Newton's Law of Inertia and the Nature of Gravitational Forces', *Philosophy of Science*, **40**, pp. 329–59.

EARMAN, J. *et al.* (*eds.*) [1977]: *Foundations of Spacetime Theories. Minnesota Studies in Philosophy of Science*, **8**, Minneapolis: University of Minnesota.

FIELD, H. [1980]: *Science without Numbers*. Oxford: Blackwell.

FRIEDMAN, M. [1974]: 'Relativity Principles, Absolute Objects and Symmetry Groups', in *Space Time and Geometry*, (*ed.* P. Suppes). Dordrecht: Reidel.

FRIEDMAN, M. [1983]: *Foundations of Spacetime Theories*. Princeton: Univ. Press. Forthcoming.

GARDNER, M. [1977]: 'Relationism and Relativity', *British Journal for Philosophy of Science*, **28**, pp. 215–33.

GEROCH, R. [1968]: 'Spinor Structure of Spacetimes', *Journal of Mathematical Physics*, **9**, pp. 1739–44.

GEROCH, R. [1972]: 'Einstein Algebras', *Communications in Mathematical Physics*, **26**, pp. 271–5.

HAWKING, S. and ELLIS, G. [1973]: *The Large-Scale Structure of Spacetime*. Cambridge: Cambridge Univ. Press.

HELLMAN, G. and THOMPSON, F. [1975]: 'Physicalism', *Journal of Philosophy*, **72**, pp. 551–64.

HELLMAN, G. and THOMPSON, F. [1977]: Physicalist Materialism, *Nous*, **11**, pp. 309–45.

HOOKER, C. [1971]: 'The Relational Doctrines of Space and Time', *British Journal for Philosophy of Science*, **22**, pp. 97–130.

HORWICH, P. [1978]: 'On the Existence of Time Space and Spacetime', *Nous*, **12**, pp. 397–419.

HUGHES, G. and CRESSWELL, M. [1972]: *An Introduction to Modal Logic*. London: Methuen.

KAPLAN, D. [1975]: 'How to Russell a Frege-Church', *Journal of Philosophy*, **72**, pp. 716–29.

KRIPKE, S. [1972]: 'Naming and Necessity', in *Semantics of natural language* (*eds*. D. Davidson and G. Harman). Dordrecht: Reidel.

LEWIS, D. [1968]: 'Counterpart Theory and Quantified Modal Logic', *Journal of Philosophy*, **65**, pp. 113–26.

LEWIS, D. [1974]: *Counterfactuals*. Oxford: Blackwell.

MACH, E. [1942]: *The Science of Mechanics* (*transl.* T. J. McCormack). La Salle Ill: Open Court.

MALAMENT, D. [1976]: 'Review of Sklar [1974]', *Journal of Philosophy*, **73**, pp. 306–23.

MALAMENT, D. [1982]: 'Newtonian Gravity, Limits and the Geometry of Space', forthcoming in a Pittsburgh collection.

MALAMENT, D. [1982*a*]: 'Review of Field [1980]', *Journal of Philosophy*, **79**, pp. 523–34.

MISNER, C. *et al.* [1973]: *Gravitation*. San Francisco: Freeman.

NEWTON, I. [1934]: *Principia*. Motte-Cajori (trans.). Berkeley: Univ. of California.

NEWTON-SMITH, W. [1980]: *The Structure of Time*. London: Routledge.

PALTER, R. (*ed.*) [1970]: *1666: The* Annus Mirabilis *of Isaac Newton*. Cambridge Mass.: MIT.

PARSONS, C. [1967]: 'Grades of Essentialism in Quantified Modal Logic', *Nous*, **1**, pp. 181 ff.

PARSONS, C. [1969]: 'Essentialism and Quantified Modal Logic', *Philosophical Review*, **70**, pp. 35–52.

QUINE, W. [1963]: *From a Logical Point of View* (2nd ed.). New York: Harper and Row.

RUSSELL, B. [1901]: 'Is Position in Time and Space Absolute or Relative', *Mind*, **10**, pp. 293–317.

SKLAR, L. [1974]: *Space, Time and Spacetime*. Berkeley: Univ. of California.

STALNAKER, R. [1976]: 'Possible Worlds', *Nous*, **10**, pp. 65–75.

STEIN, H. [1967]: 'Newtonian Spacetime', *Texas Quarterly*, **10**, pp. 174–200; reprinted in Palter, R. (*ed.*) [1970].

STEIN, H. [1977]: 'Some Philosophical Prehistory of General Relativity', in J. Earman *et al.* (*eds.*) [1977].

TARSKI, A. [1956]: *Logic, Semantics, Metamathematics*. Oxford: Clarendon Press.

VAN FRAASSEN, B. [1980]: *The Scientific Image*. Oxford: Clarendon Press.

WEINGARD, R. [1975]: 'On the Ontological Status of the Metric in General Relativity', *Journal of Philosophy*, **72**, pp. 426–31.

WHITEHEAD, A. [1920]: *The Concept of Nature*. Cambridge: Cambridge Univ. Press.

Brit. J. Phil. Sci. 35 (1984). 115–128 *Printed in Great Britain*

The Standing of Psychoanalysis

by EDWARD ERWIN

B. A. Farrell [1981] tries to elucidate some of the rational considerations that determine the standing and value of psychoanalysis. He is sceptical about much of the positive evidence, but he also tries to provide some support for Freudian doctrines. I examine his supporting arguments and try to show that they have serious weaknesses.

Much of the early philosophic debate about psychoanalysis focused on one question: Is it possible in principle to test Freudian theory? Several philosophers have now argued that it is (Wisdom [1967] Glymour [1974]; Grünbaum [1979]), and, at the moment, the issue appears to have been settled. More recently, some philosophers have begun to examine the clinical and experimental arguments in the Freudian literature (Grünbaum [1980], [1983]; Erwin [1981]; Neu [1977]). B. A. Farrell's book [1981] is likely to enhance this recent trend: it contains one of the first attempts by a philosopher to make a comprehensive assessment of the scientific standing of Freudian theory and therapy.

Farrell is sceptical about many of Freud's views, but he does argue for some important favorable conclusions. If he is right, he has not only answered Popper [1962] and Nagel [1959] on the testability issue, but also those critics (*e.g.*, Eysenck & Wilson [1973]; Wolpe [1981]; Rachman [1971]) who find no merit whatsoever in Freud's ideas.

In some of his conclusions, Farrell employs a distinction between 'High' and 'Low' level Freudian theory, one that is analogous to the more traditional distinction between the Metapsychological and Non-metapsychological hypotheses. He does not give any operational criterion for identifying the High Level hypotheses, but he does say that they give an account of the way the mind works as a whole (p. 37), and helps to explain the Low Level generalizations of the theory as well as the observable conduct of people. An example is: The mind contains mental elements, each of which has two aspects, an ideational presentation and an effective charge of energy. The Low Level theory contains at least some ground floor generalizations, such as: All young boys are sexually attracted to the mother and are sexually jealous of the father. Some Low Level generalizations, however, resemble the High Level type; they also contain theoretical concepts, such as 'repression' and 'regression,' and say something about the way the mind works. Consequently, it is not always clear exactly which hypotheses Farrell intends to include in each category. For most of his purposes, however, his distinction is clear enough.

Using the High-Low Level distinction, Farrell reaches five favourable

Received 13 *September* 1982

conclusions: (1) There is some ground for believing that some patients with neurotic difficulties improve as a result of the analysis they receive (p. 187); (2) Freudian case material lends *some* support to analytic theory (p. 140), and lends *considerable* support to some of the generalizations of the Low Level Theory (p. 141); (3) The experimental evidence supports certain parts of Freudian theory (pp. 167–8); (4) The present state of scientific play suggests that the High Level Theory is of great promise (p. 170); and (5) Psychoanalytic theory is, and will continue to be, of great heuristic value to psychologists (p. 213).

Farrell does not make extravagant claims for the *degree* of support provided for either Freudian theory or therapy; indeed, in some cases, he makes very weak claims about the epistemic value of the evidence he presents. Such evidence, moreover, is notoriously difficult to interpret; consequently, competent commentators can sometimes reasonably disagree about its cogency. Despite these reservations, it seems to me demonstrable that Farrell places too much epistemic value on much of the evidence he amasses. I shall try to show that there are serious problems with *all* of his arguments for his positive conclusions about psychoanalysis. (Farrell also reaches some favorable conclusions about non-Freudian psychodynamic theories, but I will not discuss these conclusions here.) I shall begin with the conclusion about psychoanalytic therapy.

1 In 1952 and in 1966, Eysenck argued that there was, at the time of writing, no firm evidence supporting the effectiveness of psychoanalytic therapy (or eclectic psychotherapy, which is not at issue here). In the beginning of his discussion (p. 177), Farrell correctly interprets Eysenck's conclusion as implying the absence of evidence, rather than attributing to him a much stronger conclusion (suggested by some of Eysenck's comments): that psychoanalytic therapy is ineffective. He nevertheless rejects Eysenck's argument because of Eysenck's reliance on the Landis [1938] and Denker [1946] studies to establish a 2/3 spontaneous remission rate for psychoneurotic symptoms within two years of onset. Farrell presents several cogent criticisms of these studies, but ignores other studies cited by Eysenck [1966] and Rachman [1971] that tend to support the same remission rate. The crucial point, however, is that he agrees with Eysenck that comparing a psychoanalytic success rate to spontaneous remission rates does not establish the effectiveness of Freudian therapy (p. 179). Why, then, does he say that Eysenck's argument is 'fallacious' (p. 178)? The answer is that, as his discussion develops, he changes Eysenck's conclusion from (1) We do not have the required evidence to establish the effectiveness of psychoanalytic therapy (p. 177), to the stronger conclusion (2) If we give patients psychoanalysis, they do conspicuously worse than if we give them some other form of psychotherapy or no therapy at all (p. 178). If (2) were Eysenck's conclusion, he would need solid grounds for his assumption of a 2/3 spontaneous remission rate (at least for the type of client typically

treated by psychoanalysts). To argue for (1), however, he does not need such grounds; in the absence of other kinds of evidence (such as evidence derived from controlled studies), it is sufficient to assume that spontaneous remission rates are too uncertain to be used to establish either the effectiveness or ineffectiveness of psychoanalysis. Farrell does not challenge this latter assumption; he argues for it himself and thereby helps shore up Eysenck's argument. (For a detailed defence of the argument, see Erwin [1980].)

If spontaneous remission rates cannot be used as a basis of comparison, what evidence is there for Farrell's conclusion (p. 187) that some patients improve as the result of their analysis? He cites three studies that he describes as 'fairly satisfactory' (p. 182). The first (Schjelderup [1955]) is an uncontrolled case study of 28 patients treated between 1926 and 1943. The time between the assessment of their condition and the termination of their therapy varied between eight and twenty-four years (during the hiatus some of the patients had lived through a major depression and a world war); no attempt was made to rule out the possibility that any improvement was due to events that occurred after the termination of therapy. The second study (Orgel [1958]) concerned peptic ulcer rather than neurosis (a point that is relevant if one wants to generalize the results to typical clinical problems). Orgel judged ten out of fifteen of his patients to be cured, but no control group was employed and the possibility was not ruled out that a combination of medical treatment and non-treatment factors caused the remissions. The third study was done by Cappon [1964]; it was not a study of psychoanalytic therapy. According to Cappon, the therapy was primarily Jungian, but it also included a melange of other therapies (including elements from learning theory, psychoanalysis and existentialist psychiatry). There is no way to tell from Cappon's report which treatment component, if any, was responsible for any improvement.

None of the above studies provides any evidence for the effectiveness of psychoanalysis and each has been criticized elsewhere (Erwin [1980]; Rachman and Wilson [1981]). However, Farrell himself (pp. 183–4) spells out important weaknesses of these three studies; so, I am not sure that he wants to rely on them. His remaining evidence consists of a single, uncontrolled case study of one patient treated by the late Dr P. M. Turquet. The patient's overt problem was 'intellectual under-functioning' as evidenced by his poor scholastic performance; the scholastic performance improved considerably after 2½ years of treatment. Farrell contends that we have good, but not conclusive, grounds for believing that the improvement was caused by the treatment. The only ground he cites, however, is the fact that improvement followed the therapy. How do we know that some of the other factors that Farrell himself cites—*e.g.*, maturation, school 'promotions', the client's personal relationship with his History teacher—or other factors (*e.g.*, placebo factors, the sensitivity of the therapist, etc.) did not singly or in combination cause the improvement? Without the use of controls of some kind, there is no way to identify which of a potpourri of plausible possibilities was the actual cause of improvement. A more general point should be stressed. There is a large literature explaining in great detail why uncontrolled case studies are generally too weak to provide any reliable evidence for the clinical effectiveness of a psychological therapy (*e.g.*, Scriven [1959]). Even with the use of sophisticated controls, it has proved extremely difficult to rule out placebo and other rival explanations of the therapeutic data (see, *e.g.*, Kazdin and Wilcoxin [1976]). It is surprising, therefore, that someone with Farrell's methodological sophistication would ignore all of the difficulties and rely on an uncontrolled case study of one patient to support the effectiveness of psychoanalysis. Two recent reviews of the literature (Erwin [1980]; Rachman and Wilson [1981]) conclude that there still is no evidence that psychoanalysis is an effective clinical procedure. There is nothing in Farrell's book to alter that assessment.

2 Farrell claims (p. 143) that four Low Level Freudian hypotheses receive considerable support from Freudian case studies. He acknowledges (Fenichel [1945] and Dicks [1947]) for the *formulation* of these hypotheses, but he does not say what clinical evidence, if any, confirms them. Instead he discusses some general methodological considerations concerning the use of Freudian case studies. The case material, he holds (p. 134), does not and cannot work by confirming or falsifying Freudian hypotheses, but nevertheless the theory is open to the influence of empirical facts discovered in analysis. How does this influence work? There are rational considerations of a clinico-factual kind that make the analyst strengthen his views or change them (p. 136). Farrell's main example is Bibring's [1953] modification of the classical Freudian account of the etiology of depression. As a result of his experience with patients, Bibring rejected the orthodox Freudian explanation and concluded that depression is caused by the ego being shocked into an awareness that it is helpless to realize its aspirations. Bibring, however, did not confirm his new theory; he did not present any satisfactory grounds of any kind for accepting his account of the etiology of depression. Farrell also mentions other ways in which the analyst is led to modify Freudian theory as a result of clinical observation. For example, the analyst notices features of, say, obsessional patients that have not been noticed before, or comes to see the input from the patient in a new way. All of these sorts of cases, however, present the same problem as the Bibring example. From the fact that the analyst modifies Freudian theory as a result of his clinical observations, we cannot reasonably infer that the modification is warranted. The new theory may have no more support than the old theory.

Farrell also points out that psychoanalytic narratives spun by therapists render the case material intelligible. A necessary condition for the truth of such narratives, Farrell claims, is thereby satisfied; consequently, the case material, he argues, does lend *some* support (Farrell's emphasis) to the narratives and hence to psychoanalytic theory itself (p. 129). I do not find this argument convincing. A necessary condition for the truth of any story about psychoanalytic case material, even a wildly fanciful fairy tale, is that it be logically consistent. It does not follow, however, that satisfaction of a necessary condition for the truth of such a story is *sufficient* for confirmation, or for support of any kind for the story. A psychoanalyst might try to argue that psychoanalytic narratives do more than satisfy a necessary condition of their truth; they also, at least sometimes, provide the most plausible account of the case material. Farrell, however, attempts no such argument. In fact, he argues (p. 62) that competing, non-Freudian psychodynamic theories can equally well explain the case material. In the absence of any argument to rule out competing, plausible explanations, pointing out that psychoanalytic narratives make the case material intelligible is insufficient to show that the material supports the truth of such narratives.

There are well known epistemological objections to using the case material to support psychoanalytic theory (*e.g.*, see Nagel [1959]). Farrell (pp. 47–66) repeats some of these objections and adds some useful criticisms of his own. The situation, however, is even darker than he allows: Grünbaum has recently demonstrated ([1980], [1983]) that there are even deeper and more pervasive problems in the use of Freudian case material as confirming evidence. Farrell offers no solutions to these problems, or even to the more standard difficulties that he himself raises. Consequently, he fails to show that any Freudian hypothesis is supported by clinical observations'.

3 Farrell does not try to review all of the experimental evidence bearing on Freudian theory; and some of what he cites is, he claims, either neutral or disconfirmatory. He does argue (pp. 167–8), however, that *some* of the experimental evidence supports parts of Freudian theory, although the degree of support is different for different hypotheses. In contrast, I think that the evidence he cites is seriously defective if interpreted as supporting *any* part of Freudian theory.

Several of the hypotheses Farrell discusses are not, in my view, part of Freudian theory. They are either not theoretical or they are non-Freudian. One of these hypotheses concerns the anal character. In the study cited by Farrell, Beloff [1957] quotes Freud about the origin of this character (Beloff, p. 145):

> ... The infant derives a great deal of erogenous pleasure from its bowel movements, and when this pleasure is inhibited by severe bowel training, the usual result is reaction formation. . . . In extreme cases . . . one becomes parsimonious, stingy, meticulous, punctual, tied down with petty self-restraints. Everything that is free, uncontrolled, spontaneous is dangerous . . ..

Beloff found that some of his subjects did exhibit some of the characteristics mentioned by Freud. He then proceeded to test Freud's theory that such characteristics are related to bowel training experience in infancy. He found no such evidence. In view of this failure, it would be incorrect to say that

Beloff confirmed Freud's theory of the anal character. Farrell makes no such claim. He says only that there is good ground to believe that the three traits of orderliness, parsimony and obstinacy can be found to go together (p. 165). He also points out that this cluster of traits might be an aspect of obsessionality, which has no special connection to psychoanalytic theory (p. 166).

Parallel remarks apply to the so-called 'oral' character. To support Freudian theory, it is not sufficient to show that some people have the characteristics that constitute this character; it is also necessary to confirm the Freudian explanation of the etiology or clinical significance of these characteristics. Again, Farrell apparently agrees. He points out (p. 166) that until the causal origin of the relevant trait cluster is established, it is potentially misleading to go along with analytic theory and describe them as either 'oral' or 'anal'.

Another example of a non-theoretical hypothesis is this: when male children become interested in their bodies, they first take an interest in their mouth, then their anus, and then their genitals. Farrell points out that some observations of children by Gessell (Gessell and Ilg [1942]) tended to support this generalization, but Farrell also agrees (p. 29) that the generalization does not qualify as a theoretical hypothesis; it does not contain any theoretical terms. In contrast, the Freudian explanation of why the male infant goes through these stages of interest (he is sexually driven) is theoretical, but this hypothesis, Farrell contends, runs counter to the 'state of play in science' (p. 167). So far, I think that Farrell and I agree.

I have no quarrel with the inclusion of the remaining hypotheses listed in Farrell's summary (p. 167–8) as parts of Freudian theory. One of these is the analytic account of homosexuality. Farrell does not mention any specific studies, but alludes to ones using questionnaires which ask adult homosexuals to report their present attitudes to their parents, and to recall how they were treated by their parents in their childhood. He describes such studies as providing 'weak support' for the analytic account of homosexuality (p. 168). There may be very little difference between describing the support as 'weak' or 'non-existent', but I think the latter term is more plausibly applied. It is not just that memories of childhood events are often unreliable, but rather no evidence is provided that the described events, even if they did occur, *caused* the development of the homosexuality.

The remaining two hypotheses that Farrell says are supported by the scientific evidence concern two Freudian defense mechanisms: displacement and repression. Before examining his evidence, it will be useful to look at what Freud meant by 'defence'.

In her Book, *The Ego and the Mechanisms of Defence*, Anna Freud [1946] points out that her father introduced the term 'defence' in his writings in 1894 and used it to describe the ego's struggle against painful or unendurable ideas or affects. He later abandoned the use of this term and tended to substitute the term 'repression'. Still later, in 1926, Freud reverted to the use of the old concept of defence to cover all processes that serve the same purpose: The protection of the ego against instinctual demands (A. Freud [1946], p. 46). The term 'repression' was then applied to only one of these processes.

In *Inhibitions, Symptoms and Anxiety* (S.E., [1926], **20**: pp. 87–174), Freud notes that he had substituted the term 'repression' for 'defence'; he then adds:

> It will be an undoubted advantage, I think, to revert to the old concept of "defence", provided we employ it explicitly as a general designation for all the techniques which the ego makes use of in conflicts which may lead to a neurosis, while we retain the word repression for the special method of defence which the line of approach taken by our investigations made us better acquainted with in the first instance. (S.E., [1926], **20**: p. 163).

Later in the same essay, Freud explains that different defence mechanisms may be at work in different neuroses. He concludes:

> These observations provide good enough grounds for reintroducing the old concept of "defence", which can cover all these processes that have the same purpose—namely, the protection of the ego against instinctual demands—and for subsuming repression under it as a special case. . . . It may well be that before its sharp cleavage into an ego and an id, and before the formation of a super ego, the mental apparatus makes use of different methods of defence from those which it employs after it has reached these stages of organization (S.E., [1926], **20**: p. 164).

As the above remarks make clear, Freud's concept of defence is linked to that of the Freudian mental apparatus. In particular, the concept applies to those processes having a particular purpose: the protection of the ego against instinctual demands of the Id. The term 'defence' can also be used, however, in a non-Freudian sense so that the Freudian mental apparatus is not presupposed. For example, someone who is very sensitive to criticism may be said to react 'defensively'. Terms used to describe particular Freudian defence mechanisms can, of course, also be used in a non-Freudian sense. For example, even if there is no unconscious mind and no id-ego-superego, someone may be said to 'rationalize' his past behaviour if he tries to justify it by citing a motive he knows he did not have at the time he acted. It would not follow, however, that a Freudian defence mechanism was operative. Consider one last example. Suppose we use electric shock to train two albino rats to strike each other. We then reinforce the aggressive behaviour by turning off the shock when the animals start striking each other. Finally, we replace one of the rats with a small celluloid doll. If the remaining rat strikes the doll, we can, in a loose sense of the term, describe it as 'displacing' its aggression. It would be implausible, however, to infer that there was 'displacement' in the sense that Freud used the term. We would have no reason to believe that the rat's purpose in striking the doll was to protect its ego against instinctual demands of the id.

The experiment described in the last example was performed by Neal Miller [1948]; his study is one of two cited by Farrell as evidence for

Freudian displacement. Miller (p. 176) does say, in referring to the rat striking the doll, that in psychoanalytic terms, this might be described as displacement, but he also points out that it could also be described, in stimulus-response terms, as generalisation from the other rat to the doll. Miller might mean that the stimulus-response explanation is an alternative to the Freudian explanation in this case. If he does, then I agree, although I would add that the stimulus-response explanation is clearly more parsimonious and is a good deal more plausible. Given the plausibility of this latter explanation, the behaviour of the rat is not evidence that (Freudian) displacement occurred. Miller might also mean, however, that it would be equally correct to describe the rat's behaviour as (Freudian) *displacement* or as *generalization*; that is, the two terms have the same meaning. (Miller did, of course, make a famous attempt, along with John Dollard, to translate Freudian concepts into concepts of learning theory (Dollard and Miller [1950]).) The terms, however, are clearly not synonymous. To explain a rat's behaviour in terms of either stimulus or response generalization does not imply that the purpose of the behaviour was to protect the rat's ego; use of the term "displacement" would have that implication.

The only other reference cited by Farrell to support displacement is the classic work *Frustration and Aggression* (Dollard, *et al.* [1939]). In this book, the concept of displacement is used throughout in a loose, non-Freudian sense. The authors apply it in any case where aggression directed at a given object is prevented and is then redirected towards another object (Dollard, *et al.* [1939], p. 40), regardless of why this redirection occurs. One example given is that of kicking a chair instead of one's enemy. The explanation of this behaviour might be that the ego is protecting itself from instinctual demands of the id, but, then again, it might not. This might be a case of response generalization; or, some commonsensical cognitive explanation might be correct. To assume that the kicking of the chair is automatically a case of *displacement*, no matter what its explanation, is to use the term promiscuously and in a fashion that bears little resemblance to its use in Freudian theory. Another example is that of a boy who stole a clock because he felt he had been unfairly punished. The boy was even *aware* of his motive (which rules out Freudian displacement); he was quoted as saying, 'They wouldn't let me go swimming, so I stole the clock' (p. 106). A third example is that of the rat who struck the celluloid doll. In none of the illustrations is any evidence provided for the operation of Freudian displacement, or for the activity of any other Freudian defence mechanism.

The final hypothesis considered by Farrell to have experimental support is that repression occurs. His first example of a supporting study is one by Levinger and Clark [1961] in which words regarded as emotional by the subjects evoked greater forgetting of associations than those regarded as neutral. He claims that the experiment provides support for the hypothesis that the forgetting is an indication that repression is responsible for it (p. 163). He does not give any reason for this contention, however, and I do not think it is plausible. Levinger and Clark describe their study as an attempt to demonstrate that the forgetting of word associations is related positively to their emotional significance. They were not trying to show that repression explains the connection, and they make no such claim. It is also not true that the repression hypothesis is needed to explain their results; one plausible alternative is suggested by Eysenck and Wilson ([1973], pp. 204–5).

I do not think that Farrell wants to place much weight on the Levinger and Clark experiment. He points out that it will take *very* much more than the result of this experiment (his italics) to make the concept of repression logically compulsory on us (p. 163).

Farrell's remaining evidence for repression consists of results obtained by Dixon and others on perceptual defence. He points out that Dixon was able to circumvent some of the problems inherent in earlier studies stemming from the work of Bruner and Postman [1949], and contends that 'if Dixon's work is firmly replicated, it will suggest very strongly that some internal control machinery is at work of the sort described by the theory of repression (p. 163). It should be noted that the study Farrell is referring to (Dixon [1958]) was published more than twenty years ago; if replication were likely, it probably would have occurred by now. In fact, Dixon's study has been replicated by him and his co-workers (Dixon and Haider [1961]; Dixon and Lear [1963]). The issue, then, is not replication, but whether Dixon's results or any of the other studies of perceptual defence provide evidence for repression. There is reason to be sceptical.

To say that perceptual *defence* has occurred might suggest that the perceiver reacted so as to defend against some threatening stimuli. If the defensive manoeuvre is unconscious, this suggests the operation of some Freudian-like defence mechanism. It is, then, but a short step to the conclusion that repression, or something very much like it, does occur. Some commentators did interpret the original Bruner and Postman [1949] results in this way. In the succeeding two decades, however, a debate developed involving hundreds of experiments that proved to be one of the most interesting disputes in the history of psychology. Much of the dispute centered around methodological difficulties inherent in many of the experiments, but some of it was generated by conceptual confusion. One such confusion concerned the use of the term 'perceptual defence.' As Postman ([1953], p. 298) makes clear in a later paper, the term was used by Bruner and he to describe a certain *result*: that negatively valued stimuli had, under certain conditions, higher thresholds than positively valued or neutral ones. In a typical perceptual defence experiment, stimulus words were presented to the subjects in a tachistoscope. A 'higher threshold' was defined as a greater number of tachistoscopic exposures. So, if a subject required more exposures before recognizing a negative word, such as 'raped' or 'whore', that word was said to have a higher threshold.

Postman stresses that the concept of perceptual defence was not used to explain why negative words have a higher threshold. 'It was acknowledged and, indeed, emphasized from the very beginning, that ***perceptual defence was not in itself an explanatory principle*** and that the mechanisms mediating it remained to be discovered.' (Postman [1953], p. 298, his italics). However, he also adds that Bruner and he speculated that their results were due to 'perceptual defence mechanisms' (p. 298); thus, they used the term to refer to both their results and also to the mechanism that caused the results (although the nature of the mechanism was left unspecified).

Some investigators questioned the occurrence of perceptual defence even in its 'result' sense. For example, it was argued that the subjects might be recognizing the negative words just as quickly as the positive words, but were failing to *report* their observations as quickly. Subjects might do that, for example, if they were to hesitate before using words such as 'whore' or 'raped' in an academic setting. Other investigators, who accepted the results, questioned whether any sort of *defence* was involved. For example, Postman, Bronson and Gropper [1953] tried to account for the difference in thresholds in terms of differences in familiarity between the negative, positive and neutral words. When they controlled for this difference, they failed to find higher thresholds for the negative words. In his [1961] review of the literature, Brown discusses *ten* different explanations of the results of perceptual defence experiments, and points out that many more than ten such explanations were available in the literature. The Freudian explanation of the results, then, was only one of a number of competing explanations.

In the 1960's, new methodological difficulties were discovered in perceptual defence research and much of the later research was devoted to overcoming these problems, especially the difficulty of separating out perceptual from response factors. Dixon's work did contribute to the resolution of these problems. In an excellent review of the entire controversy (Dixon [1971]), he argues, convincingly, I think, that the perceptual defence effect does occur, *i.e.*, recognition thresholds for emotional stimuli tend to be either higher or lower than for neutral stimuli. He does not claim, however, that his work, or any of the perceptual defence studies, provides evidence for repression of threatening stimuli into the subjects' unconscious. On the contrary, he denies that the literature provides any such evidence, and describes the tendency of some commentators to associate subliminal perception with an unconscious mind as 'quite irrational' (Dixon [1971], p. 227).

The main objection to the repression explanation is that perceptual defence studies provide no grounds for thinking that negative stimuli are incorporated into an unconscious mind. This objection is sometimes overlooked because many investigators, including Dixon, do speak of 'unconscious' perception. What is generally meant, however, is just that there is perception without conscious awareness. The finding of unconscious perception in this sense is neutral between saying that the negative stimuli entered one's *unconscious* or one's *preconscious* (to use Freudian categories), or even that the process was purely physiological (the account favoured by Dixon [1971]).

If we use 'unconscious' in this neutral, non-Freudian sense, and ask if there is unconscious perception and if it can affect subsequent resognition thresholds, then we are inviting a question that has been investigated by many non-Freudian theorists: Can there be learning without awareness? To answer the question affirmatively, as many behaviourists have, is not to support Freudian theory. If some sort of conditioning theory could account for all human behaviour and perception, there would be no need to invoke psychoanalytic theory at all. The difference, then, between the Freudian and non-Freudian use of the term 'unconscious' is not subtle or unimportant. The postulation of 'unconscious' perception in the non-Freudian sense is compatible with a wide range of explanations of perceptual defence effects; it is consistent with behavioural accounts (*e.g.*, Eriksen and Browne [1956]), information processing explanations (Erdelyi [1974]), and the account favoured by Dixon [1971].

To conclude, Farrell's survey of the experimental literature provides no support for any part of Freudian theory.

4 In referring to High Level Freudian Theory, Farrell writes: 'The truth is that the present state of scientific play suggests it is a story of great promise' (p. 170). If Farrell means that current scientific evidence supports, without yet establishing, important parts of the High Level theory, then I do not think he has made the case for this claim. He refers to work that indicates that the theory is congruent with the current emphasis in psychobiology on the organism as a control system and with the drift of neuropsychology, but that is true of rival theories as well. This does not provide any grounds for thinking that one of these theories in particular is correct. The only other scientific work he cites is that done by Bruner and Postman [1949] and Dixon [1958]. For reasons discussed earlier, neither of these studies provides any evidential support for Freudian theory. Farrell also points out (pp. 169–70) that the High Level theory is only temporarily untestable; it could be restated so that testable consequences could be derived from it. If his point is correct, however, it does not show that we now have supporting evidence, but only that we might someday obtain it. It may be that I have misunderstood Farrell's assessment of the High Level Theory; he may not mean that the present state of scientific play provides support for it, or parts of it. If he is not making this claim, then I may not disagree, depending on what he does mean when describing the theory as having 'great promise'.

5 Farrell also claims (p. 213) that psychoanalytic theory is, and will continue to be, of great heuristic interest and value to psychologists. In assessing this claim, it is not enough to ask if Freud's theory has had great impact. Of course it has, but we should also ask several questions about the value of that impact. First, where Freud's influence has spurred new

research, or spawned new theories or clinical techniques, have the results been fruitful? In many cases, such as research on projective tests or childhood antecedents of mental disorders or the development of neo-Freudian theories of personality or neo-Freudian clinical techniques—asking about the value of the results generates the same sorts of issues raised earlier about psychoanalysis. Is there evidence that the research leads have worked out, that the neo-Freudian hypotheses are true, or that the new techniques are effective? Second, have Freud's ideas had a negative impact? In many areas of substantial importance, I think a case can be made for saying 'yes'. For example, the acceptance of effective behavioural techniques was probably impeded for a long time because many psychologists and psychiatrists accepted the Freudian view about symptom substitution: any technique that treats so-called 'symptoms' without resolution of the client's underlying id-ego conflict will always result either in relapse or the appearance of new symptoms. Development of effective psychotropic drugs may also have been delayed because of widespread acceptance among psychiatrists of Freudian views about the etiology and maintenance of psychotic symptoms. Another example concerns the development of adequate diagnostic categories. Many of the categories in the original *Diagnostic and Statistical Manual of the American Psychiatric Association* reflected Freudian theory and were of dubious utility (Franks [1975]); their widespread use probably hampered the development of newer, more useful categories. These examples are somewhat conjectural, but not implausible; it cannot be assumed without argument that Freud's ideas have had a mainly beneficial impact. Third, even if they have, is Freudian theory still needed for heuristic purposes? A great deal more is known now about the brain, psychotropic drugs, effective behavioural treatments, etc. than was known in Freud's day. What was useful for inspiring research even twenty years ago may not be useful today. I am not saying that no case can be made for concluding that Freudian theory was and still is of heuristic value, but it is necessary to argue the case. Farrell does not do that.

Conclusion: If Farrell has made the best argument that can be made, then the grounds for accepting either Freudian therapy or any part of Freudian theory are very weak indeed. The other alternative is that a better defence can be constructed; if that is true (and I think it is), then, despite the virtues of his book, Farrell has not succeeded in settling the issue of the standing of psychoanalysis.

University of Miami

REFERENCES

BELOFF, H. [1957]: 'The Structure and Origin of the Anal Character', *Genet. Psychol. Monogr.*, 55, pp. 141–72.

BIBRING, E. [1953]: 'Mechanisms of Depression' in *Affective Disorders*, Greenacre, P. (*ed.*) International University Press.

BROWN, W. [1961]: 'Conceptions of Perceptual Defence', *Brit. J. Psychol. Monogr. Supp.* No. 35.

BRUNER, J. and POSTMAN, L. [1949]: 'Perception, Cognition and Behavior', *J. Personality*, **18**, pp. 14–31.

CAPPON, D. [1964]: 'Results of Psychotherapy', *Brit. J. of Psychiatry*, **110**, pp. 35–45.

DENKER, P. [1946]: 'Results of Treatment of Psychoneuroses by the General Practitioner', *New York State J. of Medicine*, **46**, pp. 2164–6.

DICKS, H. [1947]: *Clinical Studies in Psychopathology*. Arnold.

DIXON, N. [1958]: 'Apparent Changes in the Visual Threshold as a Function of Subliminal Stimulation', *Quar. J. Exper. Psychol.*, **10**, pp. 211–5.

DIXON, N. [1971]: *Subliminal Perception: The Nature of a Controversy*. McGraw Hill.

DIXON, N. and HAIDER, M. [1961]: 'Changes in the Visual Threshold as a Function of Subception', *Quar. J. Exp. Psychol.*, **13**, pp. 229–35.

DIXON, N. and LEAR, T. [1963]: 'Electroencephalograph Correlates of Threshold Regulation', *Nature*, **198**, pp. 870–2.

DOLLARD, J. *et al.* [1939]: *Frustration and Agression*. Yale.

DOLLARD, J. and MILLER, N. [1950]: *Personality and Psychotherapy*. McGraw Hill.

ERDELYI, M. [1974]: 'A New Look at the New Look: Perceptual Defense and Vigilance', *Psychol. Review*, **81**, pp. 1–25.

ERIKSEN, C. and BROWNE, C. [1956]: 'An Experimental and Theoretical Analysis of Perceptual Defense', *J. Abnorm. (Soc.) Psychol.*, **52**, pp. 224–30.

ERWIN, E. [1980]: 'Psychoanalytic Therapy: The Eysenck Argument', *Amer. Psychol.*, **35**, pp. 435–43.

ERWIN, E. [1981]: 'The Truth About Psychoanalysis', *Journal of Philosophy*, LXXVIII, pp. 549–60.

EYSENCK, H. J. [1952]: 'The Effects of Psychotherapy: An Evaluation', *J. of Consulting Psychol.*, **16**, pp. 319–24.

EYSENCK, H. J. [1966]: *The Effects of Psychotherapy*. Inter-Science Press.

EYSENCK, H. J. and WILSON, G. [1973]: *The Experimental Study of Freudian Theories*. Methuen.

FARRELL, B. A. [1981]: *The Standing of Psychoanalysis*. Oxford University Press.

FENICHEL, O. [1945]: *The Psychoanalytic Theory ofNeurosis*. Routledge, Kegan Paul.

FRANKS, J. [1975]: *Psychiatric Diagnosis: A Review of Research*. Pergamon Press.

FREUD, A. [1946]: *The Ego and the Mechanisms of Defence*. International Universities Press.

FREUD, S. [1926]: 'Inhibitions, Symptoms and Anxiety', in *Standard Edition of the Complete Psychological Works of Sigmund Freud, trans.* J. Strachey *et al.* Hogarth Press, 1959, volume **20**.

GESELL, A. and ILG, F. [1942]: *Infant and Child in the Culture of Today*. Hamish Hamilton.

GLYMOUR, C. [1974]: 'Freud, Kepler and the Clinical Evidence', in R. Wollheim (*ed.*) *Freud: A Collection of Critical Essays*. Anchor Books.

GRÜNBAUM, A. [1979]: 'Is Freudian Psychoanalytic Theory Pseudo-Scientific by Karl Popper's Criterion of Demarcation?', *Amer. Phil. Quarterly*, **16**, pp. 131–1.

GRÜNBAUM, A. [1980]: 'Epistemological Liabilities of the Clinical Appraisal of Psychoanalytic Theory', *Nous*, XIV, pp. 307–86.

GRÜNBAUM, A. [1983]: *The Foundations of Psychoanalysis: A Philosophical Critique*. University of California Press, forthcoming.

KAZDIN, A. and WILCOXIN, L. [1976]: 'Systematic Desensitization and Nonspecific Treatment Effects: A Methodological Evaluation', *Psychol. Bulletin*, **83**, pp. 729–58.

LANDIS, C. [1938]: 'Statistical Evaluation of Psychotherapeutic Methods', in S. E. Himie (*ed.*) *Concepts and Problems of Psychotherapy*. Heinemann.

LEVINGER, G. and CLARK, J. [1961]: 'Emotional Factors in the Forgetting of Word Associations', *Jour. of Abnormal and Social Psychol.*, **62**, pp. 99–105.

MILLER, N. [1948]: 'Theory and Experiment Relating Psychoanalytic Displacement to Stimulus-Response Generalization', *J. Abnorm. Soc. Psychol.*, **43**, pp. 155–73.

NAGEL, E. [1959]: 'Methodological Issues in Psychoanalytic Theory' in *Psychoanalysis, Scientific Method and Philosophy*, S. Hook (*ed.*) New York University Press.

NEU, J. [1977]: *Emotion, Thought and Therapy*. University of California Press.

ORGEL, S. [1958]: 'Effect of Psychoanalysis on Peptic Ulcer', *Psychosomatic Medicine*, **20**, pp. 117–23.

POPPER, K. [1962]: *Conjectures and Refutations*. Basic Books.

POSTMAN, L. [1953]: 'On the Problem of Perceptual Defense', *Psychological Review*, **60**, pp. 298–306.
POSTMAN, L., BRONSON, W. and GROPPER, G. [1953]: "Is There a Mechanism of Perceptual Defense?', *Journal of Abnormal and Social Psychology*, **48**, pp. 215–23.
RACHMAN, S. [1971]: *The Effects of Psychotherapy*. Pergamon Press.
RACHMAN, S. and WILSON, G. T. [1980]: *The Effects of Psychological Therapy*. Pergamon Press.
SCHJELDERUP, H. [1955]: 'Lasting Effects of Psychoanalytic Treatment', *Psychiatry*, **18**, pp. 109–33.
SCRIVEN, M. [1959]: 'The Experimental Investigation of Psychoanalysis', in Sidney Hook (*ed.*) *Psychoanalysis, Scientific Method and Philosophy*. New York University Press.
WISDOM, J. O. [1967]: 'Testing an Interpretation Within a Session', *The International Journal of Psycho-Analysis*, **48**, pp. 44–52.
WOLPE, J. [1981]: 'Behavior Therapy Versus Psychoanalysis', *American Psychologist*, **36**, pp. 159–64.

Brit. J. Phil. Sci. 35 (1984), 129–147 *Printed in Great Britain*

Meno—a Cognitive Psychological View

by BENNY SHANON

1 *Introduction*
2 *Meno's syllogism: Premises and supporting tenets*
3 *The underlying cognitive assumptions*
4 *The matter-like view of mind: Summary and suggestion for an alternative*
5 *Postscript: Methodological considerations*

1 INTRODUCTION

Plato's *Meno*[1,2] presents one of the most famous arguments for nativism in the history of philosophy. In this dialogue the philosopher claims that knowledge is not acquired by learning, but that it is rather represented in the mind in an innate fashion. Teaching, it follows, is not the passing of information from one person to another, but rather the bringing out of memories that the learner had already known prior to his encounter with his teacher. While Plato's stated conclusion might strike the contemporary reader as extreme and old-fashioned, the different assumptions on which Meno's argument is based, I find, are still, in one way or another, much with us.

The present paper sets itself to define the tenets on which Meno's syllogistic argument is based and to uncover the more fundamental assumptions on which they are couched. These assumptions are examined from a modern cognitive-psychological perspective which suggests that none of the tenets necessarily holds, hence that Plato's conclusion is unwarranted. Platonic exegesis, however, is not the subject matter of this discussion and it is not the determination of the particulars of Plato's actual beliefs which is of main concern to me here. Rather, my interest lies in the cognitive meta-theory which seems to be defined by the totality of different tenets and assumptions to be noted. This meta-theory, to which I shall refer as *the matter-like view of mind*, constitutes, I believe, the conceptual framework for much of contemporary cognitive research. This view is in general not explicitly acknowledged in the literature and Plato's dialogue

---

[1] The present analysis of Meno is based on the study of the original Greek text, and the following three translations: English (Plato [1937]), Hebrew (Plato [1965]) and French (Platon [1949]). Quotations are based on Jouwett's English translations, except for one place (marked with an asterisk in the text) in which a check of the original Greek revealed that the translation is not accurate. I am grateful to Dr Anita Mittwoch for her help with the original text.

[2] As used in this paper, the terms 'Meno', 'the dialogue', 'Plato's syllogism' and the like are taken in reference to sections 82–5 of the work in question. The discussion, note, is strictly cognitive, and it takes no interest in those aspects of Plato's theory which have to do with the soul and its eternality.

Received 26 *November* 1982

serves as a convenient test case to expose it and to clarify its inadequacies. In the following discussion the view will be critically examined and outlines for an alternative to it will be suggested.

2 MENO'S SYLLOGISM: PREMISES AND SUPPORTING TENETS

Meno's main syllogism runs as follows:

(*) (a) At the beginning of the session the slave does not know *X*.
(b) At the end of the session the slave knows *X*.
(c) During the session Socrates did not teach *X* to the slave, that is, never during the session did Socrates present the slave with content *X*, and assert that it is the case.
(d) Given the information *X* was not made available by any other external source, the slave must have obtained it from an internal one; in other words, he must have remembered it.

For conclusion (d) to follow from the premises (a)–(c) four supporting tenets have to be invoked. The first of these tenets is (I):

I Information may not be generated.

In other words, the change from the state described in (a) to the one described in (b) may come about in only one of two ways:

I′ Either the information *X* is introduced externally, or else it is retrieved internally, as it is, from one's memory store.

That tenet (I) is essential for (*) to pass through is evident. Without this tenet one could not have stepped from (c) through the first clause of (d) to the concluding clauses of the syllogism. Specifically, without (I) the lack of teaching could not serve as evidence from which the retrieval of information from a permanent, albeit dormant, store could be concluded. Clearly, for the negation of one alternative to imply a second alternative, the exhaustive disjunction of the two alternatives has to be postulated. Plato is fully conscious of this, for in the dialogue Socrates actually says:

> Without anyone teaching him he will recover his knowledge for himself

and

> This knowledge which he now has (he) must . . . either have acquired or always possessed.

Plato's second tenet may also be phrased as a disjunction:

II Either one knows (a given information *X*) or one does not.

In other words:

II′ Partial knowledge is not possible.

The dichotomous nature of knowledge is consistently assumed in the very practice of Socrates' encounter with the slave. On the one hand, Socrates presents all his statements as ones which, in principle, are either true or false; on the other hand, the slave is expected either to accept these statements or to reject them. Consequently, if the slave's evaluation is consonant with the statement's veridical truth value, he is right and thus in possession of true knowledge; conversely, if the slave's evaluation is not consonant in this manner, the slave is wrong and ignorant. The critical significance of (II) manifests itself in two respects all along the dialogue. First, all cognitive states (knowing, remembering, having taught, having learnt) are regarded as either being the case or not; second, they are all viewed as pertaining to a given quantity of information X. Indeed, were this not the case, the conclusion of Meno's syllogism need not have followed; this is demonstrated by (**), in which the exclusion of partial knowledge is not assumed:

(**) (a) At the beginning of the session the slave does not know X, but he knows x, where—on a certain measure of informativeness—$x < X$.
(b) At the end of the session the slave knows X.
(c) During the session Socrates did not teach X to the slave, but he did teach him y, such that—in some sense—x and y is tantamount to X.
(d) Hence, Socrates' teaching brought the slave from the state of partial knowledge to that of full knowledge of X.

Thus far we have ignored the difference in point of view which is noted in (*). Premises (a), (b) and the conclusion (d) take the point of view of the slave; by contrast, premise (c) takes that of Socrates. For (d) to follow from its premises, a relationship has to be established between the two points of view. This relationship is defined by tenet (III):

III Teaching and learning are symmetric.

Were (III) not the case, the following state of affairs could have been the case:

(***) (a) At the beginning of the session the slave does not know anything (of relevance to the topic of the conversation).
(b) At the end of the session the slave does not know X (he might know x, though).
Yet
(c) During the session Socrates presented the slave with information X (or even with x, such that—on a certain measure of informativeness—$x > X$) and asserted that it is the case.
In other words,
(d) Even though information was 'taught' to the slave, he did not 'learn' it.

A state of affairs such as the one depicted in (***), however, is not considered by Plato. Throughout the dialogue, the object of the (denied) teaching is always taken to be identical to the object of the learning and knowing, namely the full relevant information X at hand. This symmetrical identification is, in turn, founded on the view that the teaching-learning process consists (as noted in (c) of (*)) of the passing of particular contentual information from one person to another. Specifically,

III′a ‘A teaches X to B’ is equivalent to A’s passing X to B.

and

III′b ‘B learns X from A’ is equivalent to B’s receiving X from A.

hence

III′c If A teaches X to B, B learns X from A.

That Plato does, indeed, hold this view is shown by assertions like the following, actually made by Socrates in the text:

> Do you observe, Meno, that I am not teaching the boy anything, but only asking him questions.

and

> I shall only ask him, and not teach him, and he shall share the enquiry with me: and do you watch and see if you find me telling or explaining anything to him, instead of eliciting his opinion.

For the syllogism (*) to be valid still another tenet has to be invoked. Unlike (I), (II) and (III), (IV) does not make a substantive claim, one which is specifically cognitive, but rather, one which is methodological and very general:

IV One’s overt performance (and in particular, one’s verbal behaviour) is a genuine indicator of one’s knowledge and is sufficient to establish what knowledge one has.

hence,

IV′a If one knows one performs adequately.
IV′b If one performs adequately one knows.

That in Meno tenet (IV) is essential is evident. The only data Socrates uses in order to establish the slave’s state of knowledge consist of the slave’s verbal reports, all of which are taken as explicit indications as to whether the slave believes in the truth of given propositions or does not; otherwise, no dependent variables are examined. Were tenet (IV) not taken for granted, characterisations (a) and (b) could not be established as facts. In particular, it could be the case that the slave knows (either fully or partially) yet exhibits at the beginning of the session a state of ignorance; similarly, the slave could come to accomplish at the end of the session a performance which is, from an operational point of view, fully adequate, yet he could do so without knowing—for instance, by being tuned to demand characteristics or by guessing.

The four tenets we have noted are not unalike. All of them exhibit some categorical, all-or-none, pattern: either one knows or one does not, either one receives information from an external source or one retrieves it from an internal one, either one learns what one is being taught or one does not learn at all, and either one knows and performs or one neither knows nor performs. In the sequel, I will try to show that, indeed, the four tenets are all manifestations of the matter-like view of mind. By way of uncovering this view, I shall examine the further assumptions upon which each of the four tenets is based and the ramifications which they entail.

3 THE UNDERLYING COGNITIVE ASSUMPTIONS

Consider (I): On the face of it, the tenet appears to be quite innocuous: If nothing can be created *ex nihilo*, there is no reason why information should. Straightforward as it may seem, this analogy, I believe, bears serious, far-reaching ramifications, both in the context of Meno and for the study of cognition in general. Thus, it seems that the analogy is based on the hidden assumption (I.1):

I.1 Information is matter-like.

This assumption, or rather axiom, is the cornerstone of the matter-like view. It is tantamount to saying that:

I.1′ The rules governing information are analogous to the rules that govern the behaviour of material things.

(I.1′) however, does not seem to be true. For one thing, the quantity of matter in the world is constant, whereas the quantity of information in it accrues. (I permit myself, for the sake of simplicity and without any real cost in this context, to be Newtonian.)

That (I.1) is the basis for the matter-like view of the mind is not immediately apparent: The topic of (I.1) is information whereas the topic of a theory of mind is knowledge and belief. The two topics are categorically distinct. Following Popper [1972], I assume that information pertains to an idealised domain in which the structure of conceptual systems is defined in a formal, theoretical fashion, one which does not take into consideration human beings and their cognition. This domain includes formal logic and the mathematical theory of information, and, according to some views, all of mathematics, as well as the theory of linguistic competence (for a realistic characterisation of the status of this theory, see Katz [1981]); clearly, Plato's theory of ideas also pertains to this domain. Cognition, however, does not deal with idealised forms. Rather, it is concerned with information as it is actually represented and processed by the mind. Whether the cognitive structures of knowledge and belief exhibit the formal structures defined on the idealised level is an open question. Any association of information on the one hand, and knowledge on the other hand, then, has to involve the following assumption:

I.2 (Human) knowledge is the reflection of (ideal) information.

For Plato (I.2) is, of course, not a mere particular assumption invoked in the particular context of Meno, but rather a cornerstone of his entire philosophical system. That it is essential for Meno's syllogism to pass through, however, may be still worthwhile to specify. Statements (a) and (b) of (*), note, have to do with the slave's state of knowledge, that is— with the cognitive domain, but they were evaluated operationally (*cf.* tenet IV) as if they pertained to information in the idealised sense. Further, statement (c) is based on the idealised consideration of information, but its direct implication as stated in (d) is cognitive (also see tenet III). Finally, as this very discussion demonstrates, tenet (I) is essential for the deduction (*), but it assumes (I.1), which in turn may not be relevant unless (I.2) is also taken as being the case. In the Western philosophical tradition, various critiques of (I.2) have been raised, the most important of which is Kant's [1781]. Thus, consider (****):

(****) If *A* entails *B*, then if one knows *A* and one knows that *A* entails *B* one also knows *B*.

For Plato, who maintains that *episteme* is a reflection of logic, (****) is axiomatic; by contrast, for Kant (****) is a fallacy which demonstrates the significance of the synthetic *a priori*. The basic claim Kant (and following him modern logicians (see Rescher [1968])) make is that one may be cognisant of one information content, yet be ignorant of another content which is formally equivalent to it. The independence of logical form and cognitive content is such that even tautologies may be, from a cognitive point of view, of non-null import. Indications which counter (I.2) are also noted in the realm of language. As pointed out by Fodor, Bever and Garrett [1974] the moral of much experimental psycholinguistic research is that the structures of linguistic theory have psychological reality, but its processes do not. Modern cognitive psychological research (*e.g.* Wason and Johnson-Laird [1972]) further attests to the fact that human thinking is not governed by the laws of logic. As a rule, human information processing bears on semantic and pragmatic considerations, hence it is not strictly formal. For this reason (as well as for one to be detailed below) abstract formulations like the ones postulated by the mathematical theory of information (Shannon and Weaver [1949]) are not always suitable for the characterisation of cognitive processing. Indeed, while the employment of such formulations was much in vogue in the earlier days of modern cognitive psychology (Garner [1962]; Atkinson, Bower and Crothers [1965]) today this is definitely not the case.

Let us proceed one step further. If information is matter-like, and if knowledge reflects information, then knowledge is also matter-like; such a view implies the following:

I.3 To know is to have cognitive material stored in one's head.

Assumption (I.3) is three-faced: it characterises the objects of one's knowledge, the manner in which they are stored, and the relationship between manifest knowledge (as attested by actual performance) and stored representations:

I.3′a The objects of knowledge are matter-like entities.
I.3′b Memory is a container in which the cognitive material is stored.
I.3′c The knowledge attested to by one's behaviour is reflective of what is stored in one's memory.

None of these characterizations seems to hold true. To say that the objects of knowledge are matter-like entities is to say that knowledge is the conglomeration of atomistic units which specify particular contents. This is clearly not the case. Knowledge is not a repertory of fixed declarative statements specifying what the individual takes to be facts (*cf.* Wittgenstein [1922]). Rather, knowledge consists of a rich variety of structures which may be open (rather than closed and fully specified), procedural (rather than declarative, see Winograd [1975]; Bobrow and Winograd [1977]), and formal (rather than contentual or substantive). Schemata (Bartlett [1932]) and frames (Minsky [1975]), for example, consist of semi-open skeletons which specify variables whose particular values are fixed by the particular context at hand, whereas inasmuch as it is universal, linguistic competence is by and large formal (Chomsky [1975]). Further, as suggested by Bobrow and Winograd [1977], cognitive representations may have to allow for heterogeneity in terms of both types and tokens.

Turning to (I.3′b) we note that human memory is not a passive, inert storehouse that may be defined independently of the information it contains. As pointed out by Bartlett [1932] and by many researchers since, memory is a dynamic system which orders information and modifies it so as to integrate it within the structures it has already established. Furthermore, unlike standard containers memory may not be defined by a fixed capacity, that is, by a function which relates its magnitude to that of the elements stored in it, with both magnitudes being independently defined prior to the storage. Characteristically, meaningful coding and varied interpretations increase the range of one's remembrances in a manner which defies any capacity specifications. For this reason, the characterisation of human memory in terms of the abstract formalisms of the mathematical theory of information is of no avail (Miller [1956]; Fitts and Posner [1967]). The inadequacy of (I.3′a) and (I.3′b) leads to the inadequacy of (I.3′c). The foregoing discussion implies that the contents of memory are not any fixed, fully specified, cognitive entities stored as such, but rather the specifications sufficient for the active generation of knowledge (see Norman and Bobrow [1979]). Consequently, in order to enhance the range of one's manifest performance, the incorporation of additional content is not necessary. The classical demonstration to this effect is due to De Groot [1965] who showed that the manifest increase in the memorial performance of chess-masters is achieved by way of superior interpretation and without corresponding increase in the capacity of their memories.

The implication of the foregoing discussion to Meno's argument is evident. If, as assumed by Plato, to know is to be in possession of facts, and if the slave has attained full understanding of a domain of knowledge without having been presented with the facts, then, it is concluded, he must have retrieved his knowledge from an underlying memory store. Our discussion suggests, however, that the slave might neither be presented with the final cognitive product, nor retrieve it from an unconscious compartment of the mind, and nonetheless he could process the information presented to him and reach the product by some generative processes. This state of affairs, which is suggested by the bulk of modern cognitive research, is precisely what is being denied by tenet (I).

The critique of tenet (I) already bears on those of tenets (II) and (III). Tenet (II), recall, denies the possibility of partial knowledge. If information consists of facts, and if knowledge reflects information, then one knows whatever facts are available to one. In particular, it is not possible for one not to know, yet be in possession of some form of knowledge. This conclusion is clearly in line with the methodological tenet (IV). The empirically minded student of cognition (which Socrates purports to be) defines the scope of knowledge by testing people's performance. In Meno, the slave is presented with questions and he either answers them correctly or not. If knowledge is fully reflected in overt verbal behaviour, then it follows that either the slave knows or he does not. This, however, is a simplistic view of what knowledge is. Strictly speaking one might not know that *X* but nonetheless have some stance regarding it. One may deem *X* credible, think that it makes sense or believe that it is consistent with other pieces of knowledge one maintains, and still specifically not know that *X*.

Tenet (III), again, is a result of the application of a physical-like law of conservation to the cognitive domain. If information is matter-like, then if conservation of matter holds conservation of information holds as well. Like tenets (I) and (II), however, tenet (III) is far from being self-evident. Specifically, the tenet is based on the following assumption:

> III.1 The teaching-learning process consists of the direct passing of information from one person to another.

Assumption (III.1), note, is defined by the conjunction of the two final clauses of (III'a) and (III'b). The particular way it leads to (III) is detailed as follows:

(*****) (a) III.1.
(b) I.1.
(c) The passing of information is analogous to the passing of matter (from b).
(d) Just as the quantity of matter is not altered by physical

translocations, information is not altered by cognitive ones, that is (by (a)), by teaching.

(e) III.

Assumption (III.1), however, is neither necessary nor sufficient. One may pass information to another person, yet that person may fail to appreciate it. Alternatively, in order for someone to attain a given cognitive level one may have to present him with information which is—in some independently defined sense—greater than the one he eventually obtains. Conversely, one may attain a certain cognitive level even if one is presented with a lesser amount of information than any idealised theory would predict. This last state of affairs is made possible, of course, by the generative faculty discussed in conjunction with the previous tenet. As Socrates says:

> These notions have just been stirred up in him . . . without anyone teaching him, he will recover his knowledge if he is only asked questions.

Only questions? As already noted, teaching may involve the presentation of information which is not constituted by particular contents. Indeed, the good teacher is not the one who feeds his student with facts, but the one who trains him how to think.

Finally, let us turn to tenet (IV). Taken as a methodological stance this tenet may be regarded as one particular manifestation of the claim according to which:

IV.1 Overt performance is the criterial psychological measure.

This stance may itself be interpreted in several fashions, of which two particular ones will be noted. From a technical point of view the interpretation is:

IV.1′a In order to judge attainment of a particular psychological state, it is both necessary and sufficient to test overt performance.

This strategy is the one actually adopted in Meno, in which the psychological state is 'knowledge' and the overt performance is 'verbal report'. Viewed from a much more general perspective, however, the interpretation of (IV.1) is:

IV.1′b Overt behaviour is the only subject matter of psychology.

This is, of course, the basic credo of behaviourism. It is not necessary to elaborate here the various arguments that have been raised in the last two decades against psychological behaviourism in general, and against behaviouristic analyses of language in particular (see, for example, Chomsky [1959], Fodor [1968]). More specific criticism of the particular tests conducted in Meno are detailed in the methodological discussion which concludes this article. These various criticisms notwithstanding, I would like to give Plato a special credit for his allowance of memories of which one (in Meno, the slave) is not cognisant. These memories are latent, hence they do not constitute knowledge, for (and hence) no overt performance corresponds to them. In other words, while Plato seems to ignore the possibility of partial knowledge, he establishes the possibility of different types of knowledge: knowledge that one knows, and which one does not. With all his methodological behaviourism, then, Plato may be regarded as one of the forerunners of the unconscious.

Tenet (IV) may be further interpreted in a way which specifically relates to verbal behaviour:

IV.2 One's overt behaviour is a direct expression of one's thoughts.

Statement (IV.2) is, of course, false. The mere fact that lying is possible is a simple attestation to this effect. Statement (IV.2), however, may be interpreted in another fashion, so as to define a non-trivial philosophical stance:

IV.3 In sincere communication, language is the direct expression of one's thoughts.

Other phrasings of this thesis are also possible:

IV.3′a For linguistic utterances to be meaningful they have to reflective of underlying thoughts.
IV.3′b Language consists of the translation of thoughts into words.
IV.3′c Language is the reflection of thought.

As simple and sensible as these theses seem to be, their truth is not self-evident. Indeed, following Wittgenstein [1953], much of 20th century linguistic research may be regarded as demonstrative that the stance expressed by these theses is wrong. The theses may be objected to in two ways, one weak and the other one strong. The weak objection focuses on the particular relationship that holds between the linguistic expressions and the cognitive entities which are assumed to underlie them. It is noted that the change of modality is not free: language is not transparent with respect to thought and the relationship between the two is not that of complete isomorphism. Rather, language is a cognitive module in its own right and it is characterised by its own structure, operations and constraints—all of which may not be reduced and explained by general principles of psychology and behaviour. Indeed, the one tenet which is consistently noted in the various versions of the Chomskian theory is precisely this: that the mappings between the structures which map phonological representations are not isomorphic to those which map semantic ones (*cf.* Chomsky [1975], [1980]). The second objection denies the very postulation of underlying mental structures. Language should not be characterised as the translation of thought into words not because the translation is not direct, but because it is not clear whether there are any independent thoughts which may serve as the basis for such a translation. There are pragmatic situations in which one cannot distinguish between the linguistic expression and the assumed message it is supposed to convey. The import of language, one would consequently note, is not to convey but rather to render an effect: language performs, it does not necessarily express (*cf.* Austin [1962], Searle [1969]).

Whatever interpretation of (IV.3) one chooses, it seems that in all cases it is couched in the still more fundamental assumption (IV.4):

IV.4 Language is the expression of knowledge.

Given, however, that by (I.2) knowledge is the reflection of information, and by (I.1) information is analogous to matter, it follows that linguistic behaviour is analogous to the behaviour of entities in the physical world. In particular:

IV.4′a Communication is the passing of semantic entities from one person to another.

IV.4′b Verbal expression consists of the generation of such entities,

and

IV.4′c Comprehension consists of the reception of these entities and their incorporation into one's cognitive system.

Given that, by (I.3), the cognitive system is a container, the materialistic analogy is complete.

Also complete is our survey of the different assumptions that underlie the tenets on which the skeletal syllogism of Meno is based. As made clear throughout the foregoing discussion, the assumptions are all highly interrelated. As noted, the tenets are the different manifestations of one meta-cognitive theory which is both comprehensive and consistent. It is comprehensive in that it commits itself to positions regarding a wide range of cognitive issues, both substantive and methodological; it is consistent in that the characteristics of the different parts of the cognitive machinery it specifies all exhibit similar principles and modes of operation.

4 THE MATTER-LIKE VIEW OF MIND: SUMMARY AND SUGGESTION FOR AN ALTERNATIVE

By way of summary and review I shall now list the basic posits of the matter-like view of mind. In so doing, I would like to point out that this view is not an archaic Platonic ideology. While, as noted in the foregoing discussion, modern cognitive research offers diverse arguments and demonstrations that counter the different materialistic posits, the view as a whole is not at all obsolete.

First on the list is a characterisation of memory:

1. *Memory* is a container of given capacity in which *cognitive* entities are stored.

This posit underlines much of modern psychological research. In particular, it is fundamental to the Box model, which until quite recently has dominated cognitive thinking (for exposition of the model see Atkinson and Shiffrin [1968]).

2. *Cognitive* entities are material-like. In particular they are contentual, substantive and closed. There are atomistic cognitive entities which are independent of each other and from which the various complex entities are composed. All cognitive entities are, consequently, homogeneous.

I do not know whether anyone today would adhere to this posit in its entirety, but adherence to some of the claims noted is, for sure, current. One model of representation based on atomic primitives and their composition is Schank's theory of conceptual dependency (Schank [1972]).

3. *Basic cognitive states* are defined in terms of whatever cognitive material is in memory, and basic *cognitive operations* are defined as the transitions between the states. In particular, knowledge is defined by the filling in of memory with cognitive content, *learning* consists of increasing this content, and *teaching* is the passing of such content from one person to another (Schank [1972]).
4. *Knowledge* is the totality of one's *mental representations*, that is the totality of cognitive entities in one's memory store. Specifically, the repertory of capabilities attested to by one's manifest performance is fully isomorphic with what is represented in one's mind. In other words, if one performs in line with a rule *X*, say, it is concluded that *X* is mentally represented in one's mind. This rationale, note, is fundamental to Chomsky's arguments regarding the psychological reality of grammar, and hence to his argument regarding nativism (Chomsky [1965], [1972], [1975], [1980]; for a general presentation of the representational view of mind, see Fodor [1981], Pylyshyn [1980], as well as readings in Haugeland [1981]).
5. *Language* is the verbal reflection of underlying cognitive states. The paradigmatic, and most explicit, definition of the posit, and with it the matter-like view in general, is found in Wittgenstein's *Tractatus* [1922]. Implicitly, however, it seems to me that the posit is taken by practically all of modern psychological research to be true (Bloomfield [1933]; Miller [1951]; Clark and Clark [1977]).

Taken together these posits define a state of affairs which may be pictorially schematised by the metaphor of a box and a package. Memory is likened to a box, knowledge is a package, cognitive dynamics consists of packaging and unpackaging, and verbal communication is the transfer of a package between the boxes pertaining to different individuals.

In contrast to this box view an alternative interactional meta-theory may be put forward, the basic posits of which are the following:

1′. Memory is an active, integrative system. It is not an inert container which stores information (if it stores anything at all) but rather it is a complex dynamic system which is best characterised not by capacity but rather by the action of its operations (see for instance Bartlett [1932]; De Groot [1965]).

2′. Cognitive entities are heterogeneous, in both types and tokens. As noted, they may be procedural, formal and open (see Bobrow and Winograd [1977]), and they may not be reduced to primitives of which they are composed in any additive fashion (for a recent criticism of the primitivistic analyses, see Fodor, Garrett, Walker and Parkes [1980]; Winograd [1980]; as well as Kripke [1972] and Putnam [1975]).

3′. Cognitive operations involve the multifarious interactions between different cognitive modules whose function is not only to process the contents of given semantic representations but also to generate new representations on the basis of old ones (Norman and Rumelhart [1975]).

4′. Knowledge (as attested to by one's performance) is not isomorphic to what is stored in one's memory. Given the nature of cognitive entities and operations, performance need not be a direct reflection of mental representation. As attested by De Groot's [1965] studies, increased performance need not require a parallel increase in one's mental representations. Thus, knowing is not tantamount to having information specified in one's mind, but is rather the ability to generate behaviour which, to an external observer, appears as the manifestation of knowledge.

Correlative to this last characterisation of knowledge is the following characterisation of language:

5′. Language need not reflect underlying cognitive states: words are not the externalisation of propositions (the originator of this view is later Wittgenstein [1953], [1964]; for new lines of thinking along this vein the reader is referred to the theory of autopoiesis as expounded in Maturana [1978], Varela [1979], Maturana and Varela [1980] and Flores and Winograd, in press).

5 POSTSCRIPT: METHODOLOGICAL CONSIDERATIONS

I could have stopped here, but let me go further and continue with the analysis of Meno and consider it as if it were a psychological report. Such a perspective is not entirely foreign to the dialogue. Plato himself presents his argument in the setting of an empirical demonstration. In fact, what he describes is one of the oldest records of psychological experimentation in our culture. That the experiment may not have been carried out is utterly irrelevant; what is important rather is the principle that an experiment is viewed as a useful vehicle for the further validation of a cognitive claim. Furthermore, the construction of the dialogue as if it were the report of an experiment is in itself of psychological interest. Specifically, the report describes methodological procedures which the author, even if he did not follow them in practice, seems to consider valid. Given this perspective, the final section of our discussion will address itself to the examination of the methodological procedures of Meno. The examination will demonstrate that Plato's procedures are seriously flawed. Such a demonstration is, I believe, interesting in two respects: first, it implies that despite his explicit denial of the fact, Socrates does pass direct information to the slave (in other words, he does teach him in the sense of III); second, it points to numerous biasing rhetorical devices which are of intrinsic linguistic significance and which are also relevant to current psychological research (*cf.* Loftus [1975]).

Let us look at some numbers. There are 42 queries in the dialogue; of these 13 are not associated with knowledge and belief, but are rather rhetorical and they relate to the identification of drawings (or parts thereof) and of their demarcation. Of the 29 informational queries that remain, 17 are yes-no questions, 2 are Wh-questions which are immediately (that is, before the slave is given the opportunity to respond) transformed into yes-no ones, and 8 are Wh-ones; another 2 questions have to do with the progress of the conversation; in one, Socrates asks the slave whether he still remembers one piece of information, and in the other the end of the session is marked.

The distribution noted exhibits a clear bias in favor of yes-no questions at the expense of Wh-ones. This bias is worth commenting upon. From a methodological point of view each of the two types of questions has its own merits as well as shortcomings. Wh-questions impose minimal bias on the assessment of the informant's cognitive repertoire: they let the informant respond freely, they do not offer the facilitation of familiarity or recognition, and they leave relatively little space for guessing. The data these questions generate, however, is varied, and may not always be of the type in which the experimenter is interested. Yes-no questions, by contrast, are focal. They are useful for the checking of particular matters of fact and they generate data which allow easy codification and manipulation. The present context, if examined as a psychological experiment, is clearly one in which Wh-questions are more suitable. In Meno, Socrates is not interested in the testing of one particular model (*e.g.* one of information processing), but rather he wants to delineate what the slave knows and what he does not. In light of this aim, the marked bias in favour of yes-no questions is curious. In the sequel we shall note that not only the frequency of these questions, but also their particular phrasing and use, do cast grave reservations regarding the appropriateness of Socrates' conduct in the session, and of the validity of Plato's argumentation which is based on it.

Thus, let us consider the yes-no question in more detail. Experimental psychologists have been aware of the biasing danger associated with these questions, and have perfected various precautions for their use. The precautions are rudimentary in the design of behavioural research. First, the questions should be phrased in such a way that if the subject were to respond correctly to all of them he would say 'yes' as many times as he would say 'no'. Second, the questions' surface form should be balanced: some (ideally, half) should be in the affirmative mode, and others (ideally, half) should be in the negative one. Finally, (ideally) the two variations should be orthogonal; in other words: the mode in which the question is posed should be independent from the truth value of the answer which corresponds to it. Does Socrates adhere to these precautions? Of the 19 yes-no questions he poses, only two (!) call for a (correct) negative answer and one is phrased in the negative. A direct examination of these questions will reveal the seriousness of this effect. The question phrased in the negative is a leading one, and raises issues that will be taken up presently. Of the two questions calling for a negative answer one is tautological:

And four times is not double?

whereas the second is answered incorrectly by the boy. The erring is telling, for to no other yes-no question does the boy furnish an incorrect answer. The yes-no questions, we conclude, do, indeed, exercise a significant bias on the way the slave answers. It is questionable, then, whether they do, in fact, check his knowledge or ignorance.

It turns out that the questions, both yes-no and Wh-ones, are even more problematic. Let us start with the questions of the first type. Of the 19 questions noted only 2 are free from any hints. The others are phrased in a manner which clearly indicates what the correct answer should be. These two questions are characterised by the fact that neither addresses itself to facts or beliefs which are particular to the subject matter at hand. One question is a simple arithmetical one, whereas the other one consists of the check of evident possibilities, and as such may be trivially answered in the affirmative:

A square may be of any size?

All other questions are phrased in a manner which makes it clear what the expected—correct—answer is. The biased phrasing is of one of two types. In 80% of the cases the question contains linguistic function words which mark the truth value of the expected answer. The following examples are representative and their import is, I presume, self-evident; in these examples the hints are italicized:

And is not that four times four?
And are there not these four divisions in the figure, each of which is equal to the figure of four feet?
And if one side of the figure be of two feet, and the other side of two feet, how much will the whole be? *Let me explain*: if in one direction the space was of two feet, and in the other direction of one foot, *would the whole be anything other than* two feet taken once?*

The last question, note, opens with a Wh-form and immediately transforms itself into a leading yes-no question.

The second type of biased phrasing, the one encountered in all the remaining yes-no questions in the dialogue, is exemplified by the following:

---

* See footnote 1 above.

And *do you know that* a square figure has these four lines equal?

This phrasing is interesting, for practically speaking it leaves the respondent no option but to answer in the affirmative. The reasons for this are two-fold. First, there are logical considerations. The verb 'know' is one which introduces its complement as a presupposition; in other words (see Keenan [1971]), the complement is invariant under negation. Thus, strictly speaking, when one negates a statement such as 'I know that *X*', that is, when one utters: 'No, I do not know that *X*', one is still left with the presupposition that *X* is the case. Thus, questions introduced by *do you know that* are misleading. On the face of it, the questions seem to be concerned with a propositional attitude. In answering the question, however, one cannot confine oneself to the attitude alone, and, despite oneself, one is forced to take a stance *vis-à-vis* the contentual statement which is presupposed by the question. The foregoing considerations are theoretical; in practice, it is still possible for the speaker to respond with a negative statement, but then he is obliged to make it clear that the presupposition of *X* is cancelled. For example, he may address his interlocutor and say:

I realize that you take *X* as true, but I don't know whether this is in fact the case.

It is there that the considerations of the second kind are relevant. The considerations are sociological. The conversational situation in Meno is, clearly, not symmetric: Socrates is superior and the slave inferior. Socrates is, as far as the matter at hand is concerned, omniscient, whereas the slave is ignorant. Thus, *a priori*, the optimal strategy the slave could adopt is to follow the hints of his master. As the slave has no evidence against the veracity of *X*, there is really no reason whatsoever for the slave to adopt any other course of action, but to consent. Taking together the semantic and pragmatic considerations noted, the question form at hand may, in effect, be phrased as:

I tell you *X* is true. Do you know that *X*?

In this form, it is clear that the question not only biases the informant's response, but it also transmits information. In other words, it consists of the very teaching whose existence Plato denies.

Still, there are Wh-questions in the dialogue. In response to these, it turns out, the slave does not fare as well as he does with the yes-no (and leading) questions. Of the slave's 8 answers to Wh-questions, two are wrong, two are ignorant ('I don't know') and the four which he answers correctly are simple arithmetical ones such as:

And how many are twice two feet? Count and tell me.

The slave answers 'Four', but there is no reason to suppose that he did not know that before; in other words, with respect to this particular information assumption (a) of (*) does not hold.

In sum, the methodological criticism outlined here casts a two-fold shadow. First, it appears that Meno cannot serve as an empirical demonstration for the thesis it purports to defend. Indeed, and this is the second conclusion to be drawn from the criticism, Socrates' practice is definitely not in line with what he claims that he is doing. He claims that he does not pass direct information to the slave, but, indeed, under a very thin disguise he does. A masterly performance of a sophist, I should say.

Having criticised Meno on both theoretical and logical counts, I would like to make two positive comments. First, it should be clear that the criticism outlined here does not constitute a proof against Plato's substantive thesis of nativism. I have tried to show that the particular argumentation developed in Meno does not constitute a valid demonstration of the thesis, but this does not imply that the thesis is not valid. In itself, however, nativism is not the subject matter of the present paper. Let me just note that the phrasing of the thesis of nativism may, in turn, also be affected by the matter-like view of mind. Specifically, various formulations of the thesis may be noted: Some, like Plato's, would regard the innate cognitive component as composed of particular contents (for a modern version closest to this extreme formulation, see Fodor [1975]; others, like Chomsky [1980], would view it as defined by formal structures and constraints; others still may confine the innate component to general principles without commitment to any fixed and particularized contents or forms (see Piaget's stance in Piatelli-Palmerini [1980]). In some form or another, it appears, however, that most contemporary thinkers would subscribe to the view that a certain component of the cognitive system has to be innate.

My second comment is meant to direct the reader to the constructive lesson he might draw from our dialogue. My criticism of Meno should not imply that the dialogue is totally devoid of interest. Undoubtedly, the section of the dialogue we have examined constitutes a beautiful creation in the art of teaching. It demonstrates that a thoughtful, sensitive and considerate teacher may guide another person from a state of ignorance to one in which he acquires intellectual understanding. The teacher accomplishes this task not by passing fully articulated pieces of contentual information, but rather by drawing his pupil's attention to points of relevance, by directing his perception of things, by supplying him with general patterns of thinking and reasoning, by guiding him along difficult lines of deduction, and by giving him—when needed—helpful hints and partial information. Furthermore, Socrates' pedagogy (in the literal sense of the term) exhibits other important characteristics that the good teacher should master. He is patient, encouraging, kind and not at all menacing. In so doing not only does he attain his objective of bringing his pupil to the final state of knowledge, but he incites his interest and develops his motivation to learn. That Socrates gives all the credit to his pupil, rather than to himself, is yet another quality that demonstrates what a humane teacher he is. Turned upside down, then, Meno is not a proof that teaching is not possible, but rather a most impressive demonstration of what teaching should—and can—be.

REFERENCES

ATKINSON, J. R., BOWERS, G. H. and CROTHERS, E. J. [1965]: *An Introduction to Mathematical Learning Theory*. Wiley.

ATKINSON, R. C. and SHIFFRIN, R. M. [1968]: 'Human Memory: A Proposed System and its Control Processes.' In K. W. Spence and J. Spence (*eds*.), *The Psychology of Learning and Motivation*, vol. 2. Academic Press.

AUSTIN, J. L. [1962]: *How To Do Things with Words*. Harvard University Press.

BARTLETT, F. C. [1932]: *Remembering: A Study in Experimental and Social Psychology*. Cambridge University Press.

BLOOMFIELD, L. [1933]: *Language*. Holt.

BOBROW, D. G. and WINOGRAD, T. [1977]: 'An Overview of KRL, a Knowledge Representation Language', *Cognitive Science*, **1**, pp. 3–46.

CHOMSKY, N. [1959]: 'Review of B. F. Skinner's Verbal Behavior', *Language*, **35**, pp. 26–58.

CHOMSKY, N. [1965]: *Aspects of the Theory of Syntax*. MIT.

CHOMSKY, N. [1972]: *Language and Mind*. Harcourt Brace Jovanovich.

CHOMSKY, N. [1975]: *Reflections on Language*. Pantheon Books.

CHOMSKY, N. [1980]: *Rules and Representations*. Columbia University Press.

CLARK, H. H. and CLARK, E. V. [1977]: *Psychology and Language*. Harcourt Brace Jovanovich.

DE GROOT, A. D. [1965]: *Thought and Choice in Chess*. Mouton.

FITTS, P. M. and POSNER, M. I. [1967]: *Human Performance*. Brooks/Cole.

FLORES, C. F. and WINOGRAD, T. [in press]: *Understanding Cognition as Understanding*. Ablex Pub.

FODOR, J. A. [1968]: *Psychological Explanation: An Introduction to the Philosophy of Psychology*. Random House.

FODOR, J. A. [1975]: *Language of Thought*. T. Y. Crowell.

FODOR, J. A. [1981]: *Representations*. MIT.

FODOR, J. A., BEVER, T. G. and GARRETT, M. F. [1974]: *The Psychology of Language*. McGraw Hill.

FODOR, J. A., GARRETT, M. F., WALKER, E. C. T. and PARKES, C. H. [1980]: 'Against Definitions', *Cognition*, **8**, pp. 263–368.

GARNER, W. R. [1962]: *Uncertainty and Structure as Psychological Concepts*. Wiley.

HAUGELAND, J. [1981]: *Mind Design*. Bradford Books.

KANT, I. [1781]: *Critique of Pure Reason* (N. K. Smith (trans.)) (1953). Macmillan.

KATZ, J. J. [1981]: *Language and Other Abstract Objects*. Rowman and Littlefield.

KEENAN, E. [1971]: 'Two Kinds of Presupposition in Natural Language.' In C. Fillmore and D. T. Langendaen (*eds*.), *Studies in Linguistic Semantics*. Holt, Rinehart and Winston.

KRIPKE, S. A. [1972]: 'Naming and Necessity.' In D. Davidson and G. Harman (*eds*.), *Semantics of Natural Language*. Reidel.

LOFTUS, E. F. [1975]: 'Leading Questions and Eyewitness Report', *Cognitive Psychology*, **7**, pp. 560–72.

MATURANA, H. R. [1978]: 'Biology of Language: The Epistemology of Reality.' In G. A. Miller and E. Lenneberg (*eds*.), *Psychology and Biology of Learning and Thought: Essays in Honor of Eric Lenneberg*. Academic Press.

MATURANA, H. R. and VARELA, F. J. [1980]: *Autopoiesis and Cognition*. Reidel.

MILLER, G. A. [1951]: *Language and Communication*. McGraw Hill.

MILLER, G. A. [1956]: 'The Magical Number Seven, Plus or Minus Two: Some Limits on Our Capacity for Processing Information', *Psychological Review*, **63**, pp. 81–97.

MINSKY, M. [1975]: 'A Framework for Representing Knowledge'. In I. Winston (*ed*.), *The Psychology of Computer Vision*. McGraw Hill.

NORMAN, D. A. and BOBROW, D. G. [1979]: 'Descriptions: An Intermediate Stage in Memory Retrieval'. *Cognitive Psychology*, **11**, pp. 107–23.

NORMAN, D. A. and RUMELHART, D. E. [1975]: *Explorations in Cognition*. W. H. Freeman.

PIATELLI-PALMERINI, M. (*ed*.) [1980]: *Language and Learning: The Debate Between Jean Piaget and Noam Chomsky*. Harvard University Press.

PLATO [1937]: *The Dialogues* [B. Jouwett (trans.)]. Random House.

PLATO [1965]: *The Complete Works* [J. Libbs (*trans.*) (in Hebrew)]. Shocken.

PLATON [1949]: *Gorgias et Menon* [A. Croiset and L. Bodin (*trans.*)]. Les Belles Lettres.

POPPER, K. R. [1972]: *Objective Knowledge–An Evolutionary Approach*. Clarendon Press.

PUTNAM, H. [1975]: *Mind, Language and Reality*. Philosophical papers, Vol. 2, Cambridge University Press.

PYLYSHYN, Z. W. [1980]: 'Computation and Cognition: Issues in the Foundation of Cognitive Science', *The Behavioral and Brain Sciences*, 3, pp. 111–69.

RESCHER, N. [1968]: *Topics in Philosophical Logic*. Reidel.

SCHANK, R. C. [1972]: 'Conceptual Dependency: A Theory of Natural Language Understanding', *Cognitive Psychology*, 3, pp. 552–631.

SEARLE, J. R. [1969]: *Speech Acts: An Essay in the Philosophy of Language*. Cambridge University Press.

SHANNON, C. E. and WEAVER, W. [1949]: *The Mathematical Theory of Communication*. University of Illinois Press.

VARELA, F. J. [1979]: *Principles of Biological Autonomy*. North Holland.

WASON, P. C. and JOHNSON-LAIRD, P. N. [1972]: *Psychology of Reasoning: Structure and Content*. B. T. Batsford.

WINOGRAD, T. [1975]: 'Frame Representations and the Declarative Processes.' In D. G. Bobrow and A. Collins (*eds.*), *Representation and Understanding: Studies in Cognitive Science*. Academic Press.

WINOGRAD, T. [1980]: 'What Does it Mean to Understand Language?', *Cognitive Science*, 4, pp. 209–42.

WITTGENSTEIN, L. [1922]: *Tractatus Logico-Philosophicus*. Routledge and Kegan Paul.

WITTGENSTEIN, L. [1953]: *Philosophical Investigations*. Blackwell.

WITTGENSTEIN, L. [1964]: *The Blue Book*. Blackwell.

Brit. J. Phil. Sci. 35 (1984), 149–159 *Printed in Great Britain*

Discussions

BEHAVIOUR THERAPY: A RESPONSE TO EDWARD ERWIN

It was astonishing to come across such attempts to resuscitate the two-cultures debate as Edward Erwin's response in the March, 1982, issue of this journal to my review in the March, 1981, issue. And since his Discussion publicly impugns both my professional qualifications and my scholarly integrity, it cannot go unchallenged. I am, however, surprised to find a philosopher resorting to *ad feminam* argument. Whatever Dr Erwin's practices in the classroom may be, I myself find *ad hominem* discussion so distasteful that I do not tolerate it in my students' papers, on grounds of logic as well as ethics. But unfortunately, since Dr Erwin has chosen to engage our discussion on the personal level, he leaves me no recourse but to respond in kind, at least so far as my own background is concerned.

On that score, I grant Dr Erwin his minor premise. I was, of course, 'trained in English literature.' Why he therefore assumes, however, that a professor of English is not trained in anything else, and is necessarily unfamiliar with both psychology and the philosophy of science, to say nothing of caution, evidence, or ambiguities, is an extraordinary leap of faith. The suppressed major premise in his enthymeme is simply ridiculous. In this connection, I recommend to him the concluding section of Roger Sperry's Nobel Prize speech, delivered in Stockholm on 8 December 1981. Sperry pointed out ([1982], p. 1226) that one of the indirect consequences of his work has been the necessity for science to recognize 'the whole world of inner experience' which has traditionally been the domain of the humanities. We can, he says, no longer speak of 'the two cultures'; it has now been proved, *on scientific grounds*, that the scientific and humanistic views of human beings constitute a continuum. 'A unifying new interpretive framework emerges with far-reaching impact not only for science but for those ultimate value-belief guidelines by which mankind has tried to live and find meaning.'

As a matter of fact, I was also trained in philosophy. And for the past decade or more, a good deal of my scholarly work and publications have been in the general area of the interdisciplinary analysis of ideas, particularly at the interface between the sciences and the humanities. I am a member and have been an officer of various interdisciplinary professional associations. Moreover, I was for three years Director of The Institute for Humanistic Studies at the State University of New York at Albany, a group of research scholars (including psychologists) from a variety of disciplines, externally selected by a ten-member national Distinguished Advisory Council, who met monthly for discussions, in an interdisciplinary context, of their current research. Under my directorship, The Institute achieved international visibility and was cited favorably for its interdisciplinary work by the blue-ribbon Commission on the Humanities in its report, *The Humanities in American Life* (pp. 114–15). Although I recently stepped down from the directorship of this Institute, I am still a Senior Fellow in it. I will concede that my clinical contacts with behaviour therapy have been limited to using its techniques (successfully) to shed a cigarette-smoking habit of many years' standing. But I've read and published a good deal in the general area of psychology. And if Dr Erwin has any contacts with his colleagues in the Departments of English, French, and German, he must surely know that some of the most penetrating insights in psychology today emanate from literary theorists.

A recital of these personalia is embarrassing, and in any case the highlights of my *curriculum vitae* are readily accessible in several standard reference works (*e.g.*, *Who's Who in America*, 42nd ed.). But the nature of Dr Erwin's attacks leaves me no alternative. If I had followed his method, I could have pointed out that his knowledge of behaviour therapy seems to stem from extended study in the stacks of his university library. However, whether his knowledge stems from clinical experience or reading is irrelevant. I was reviewing his book, not assessing his skills as a clinician.

Having disposed—so far as one ever *can* dispose of *ad hominem* attacks—of Dr Erwin's personal comments, I shall turn to his criticisms of my review. In his second paragraph, he says that I 'complain' of his carefully hedged style. I do not 'complain' of it; I simply describe it, even paying it the compliment of imitation in my second sentence. And indeed Dr Erwin himself insists throughout his Discussion that nearly every statement he makes in his book *is* carefully hedged, representing not his own views but those of the authorities he cites. But this is precisely the point I make, at considerable length, in the first two paragraphs of my review. For example, I say (p. 85) that his 'consistent posture is one of precarious balance, presenting arguments on both sides of nearly every issue he examines'. Consequently I fail to see the grounds for his objection. He is of course free to write in any way he finds congenial. I will remark only that, if he has but two styles in his writing armamentarium, the careful hedging of his book is preferable to the rumbustious spleen of his Discussion. I also indicate that his methodology is to summarize findings from eighteen pages of references in what is known technically (by scientists) as a 'review of the literature' or (by humanists) as a 'handbook'. This is of course a well-recognized scholarly technique, and the care and thoroughness with which he has handled it, as I indicated in my review, are praiseworthy.

In my third paragraph I leave Erwin's style and methodology, turning to his substantive findings. But I had already entered the caveat that 'statements [in my review] which . . . sound forthright are qualified by numerous restrictions in the original'. The difference in perspective between Dr Erwin's view of his style and mine is that he continues to insist on the particular instances of it in each case, whereas I had elevated it to a general principle which, as I carefully indicated, is operative throughout his book. Indeed, my entire second paragraph is devoted to a summary of the formulae by which he introduces his restrictions, both positive and negative. Consequently Erwin's objections in his Sections §2, §3, §4, §5, and §7 are classic instances of failing to see the forest for the trees.

In paragraph 4 (p. 53) of his Discussion, Dr Erwin says that 'in quoting or paraphrasing an author's remarks, . . . minimum standards of accuracy . . . must be met or the result will be shockingly inept scholarship'. My only quibble with this statement is that I should replace 'minimum' with 'absolute'. The trouble here lies in the fact that Dr Erwin does not know the rules of quotation, which I observed scrupulously. If he had consulted almost any good handbook of composition, he could have learned them. However, I shall quote from the University of Chicago Press *Manual of Style*, which has for decades been the Bible for authors, editors, and copy-editors.

Dr Erwin's charges concern two matters: the handling of capitalization, and the handling of ellipses. On the first of these, the *Manual of Style* says (pp. 236–9): 'The initial letter [of a quoted passage] may be changed to a capital or a lowercase letter . . ., and a final punctuation mark may be changed to make the quotation fit in the syntax of the text'. More generally, 'the skill with which an author incorporates fragmentary quotations into his own text reflects his awareness of syntax, verb tenses, personal pronouns, and so forth. He will quote only as much of the source as he needs and will phrase his sentence in such a way that the quoted words fit logically, as grammatical parts of the sentence. . . . When a quotation, either run into or set off from the text, is used as a syntactical part of the author's sentence, it begins with a lowercase letter, even though the original is a complete sentence beginning with a capital. . . . The initial letter of a block quotation may also be lowercased if the syntax demands it. . . . Similarly, if a quotation is a complete sentence, even though it forms part of a sentence in the original, a lowercase letter may be changed to a capital where the construction of the text suggests it'.

On ellipses, the *Manual* (pp. 244–8) makes the following points: 'Any omission . . . must be indicated by ellipsis points. . . . Three dots indicate an omission within a sentence or between the first and last words of a quoted fragment of a sentence. . . . Four dots . . . indicate the omission of (1) the last part of the quoted sentence, (2) the first part of the next sentence, (3) a whole sentence or more, or (4) a whole paragraph or more. . . . What precedes an ellipsis indicated by four dots should be a grammatically complete sentence, either as it is quoted or in combination with the text preceding it. Similarly, what follows four dots should also be a sentence. In other words, unless the reader is to be hopelessly muddled, every succession of words preceding or following *four* ellipsis points should be functionally a sentence. The careful writer should quote enough from his source and so arrange his text that the reader will be able to grasp the sense without tripping over the syntax'. And finally, 'ellipsis points are seldom used at the beginning or end of a quoted passage. After all, unless it is the opening or closing sentence in a work that is being quoted, something precedes and follows the passage, and it is not necessary to emphasize the fact'.

Incidentally, I did not include page references because I was for the most part simply summarizing and commenting on Dr Erwin's chapters in the order in which they appeared. Anyone who had the Table of Contents before him could easily follow the argument in his book. In any event, the purpose of a review is to identify a book's potential audience and—provided the reviewer recommends the book—to persuade that audience to read it, not to set up a ping-pong debate with the author, with the reader continually checking a passage by the author against a passage by the reviewer. The reader will presumably make up his own mind, on the basis of his background, prejudices, and learning, whether he agrees with either, both, or, indeed, neither.

On the question of paraphrase, which Dr Erwin discusses in his Section §1, he clearly fails to recognize the allusion in the passage of mine that he quotes. It was to Boswell's *Life of Samuel Johnson*, and refers to an episode on 6 August 1763, in which Johnson is seeing Boswell off to Holland from Harwich:

> After we came out of the church, we stood talking for some time together of Bishop Berkeley's ingenious sophistry to prove the non-existence of matter, and that every thing in the universe is merely ideal. I observed, that though we are satisfied his doctrine is not true, it is impossible to refute it. I never shall forget the alacrity with which Johnson answered, striking his foot with mighty force against a large stone, till he rebounded from it, 'I refute it *thus*'. This was a stout exemplification of the *first truths* of *Père Bouffier*, or the *original principles* of Reid and of Beattie; without admitting which, we can no more argue in metaphysicks, than we can argue in mathematicks without axioms.

Since this is a classic refutation of metaphysical idealism, much anthologized and frequently cited, I might perhaps be forgiven for having expected a professional philosopher to recognize it. I have no wish to gloat over the lacunae in Dr Erwin's culture, but it pains me to see him flaunting the defects of the American educational system so egregiously in the pages of a prestigious British journal. However, I hope he will now see that I was, in fact, complimenting him for his tough-minded common sense over the issue of metaphysical behaviourism. There was, of course, as Dr Erwin in §1 insists, 'no reason, in talking about views of behaviour therapists, to discuss a doctrine they generally do not espouse'. I willingly concede what I have not debated.

With regard to Dr Erwin's Section §3, the issue is not quite as simple as he makes it sound, although I did not go into any detail because his book was about behaviour therapy, not about psychiatry. The real problem lies in the fact that, once an individual has been classified as a mental patient, he has taken the first step on the road that may lead ineluctably, without his acquiescence and even contrary to his expressed wishes, to electroshock and lobotomy. 'The Moral Career of the Mental Patient', by the late Erving Goffman, is the classic short study of how this happens. It is included in his *Asylums*, pp. 125–69.

In his Section §4, Dr Erwin has still failed to grasp my point, which I think is perfectly clear. In the sentences that I quoted, he uses the same words and phrases as both causal principles and as the effects that they cause. The reason for omitting the modifiers in the first sentence was to demonstrate that this was what he had done and to make obvious the circularity of his reasoning. Since his entire discussion of empiricism seemed important to me, I felt that he should 'rework his interpretation' of it in case his book 'goes into a second edition'. I was trying to be helpful, but there are none so blind as those who will not see.

Erwin takes umbrage with me in his §2 and §7, I suspect, less because of the way in which I presented his argument than because of the fact that I disagree with him. He cites Baier's review approvingly. But Baier says (p. 505) that Erwin 'makes an excellent case for saying that some of the major psychoses really are mental illnesses', citing schizophrenia as an example. However, Baier thinks that Dr Erwin has answered Thomas Szasz on this point; I do not, and explain why.

Dr Erwin's §9 is another example of a compliment which he misinterprets as a 'slur'—this one involving Leonard Krasner. There is a vast difference, when ethical choices must be made, between spinning esoteric theories in the quiet of the study and making on-the-spot decisions involving clinical practice on the front lines. Krasner's changed attitude to mental hospitals bespeaks a flexibility and open-mindedness which I commended; his very eminence made him an obvious choice to single out as representative of those behaviour therapists who I said 'seem to be making honest efforts to respond to criticism of the ethical dimensions of their work'. And this whole section was in the context of my generally approving account of Dr Erwin's treatment of the ethics of behaviour therapy. Nor did I say that Krasner himself dragged any catatonic schizophrenics to the meeting room where the therapists addressed them. I said that, as one of the co-authors of the reference cited, he 'described dragging' them. I have no more notion than does Dr Erwin as to who was physically involved in the dragging (or leading). But I did not 'guess' that it was Krasner, or indeed anyone at all. If I were to make a guess, I should surmise that it was one or more hospital orderlies. I simply cited Dr Erwin's paraphrase of Ullman and Krasner's description, although he himself mentions 'one of the authors' as addressing the catatonics (Erwin [1979], p. 62), followed three times by the word 'therapists' in the plural.

On two general issues, I fear that Dr Erwin and I will simply have to agree to disagree. In this connection, I refer him to Peter Gould, who writes (pp. 7–8):

> Such absurd oppositions as the sciences *versus* the humanities actually caricatured instead of clarifying positions. These clashes tended to obscure deeper questions—questions that even today lie buried deep in the terrain of many of the social sciences, ticking away quietly until the day of their own detonation. . . . We know now that the complex interactions of the human and physical realms are not to be illuminated simply from a single stance and a particular perspective. . . . The interpretation of a written text . . . depends upon the set of values that each person brings to it. Differences in opinion about meaning are not settled by critical experiment (as in the physical sciences), but are discussed in order to gain the assent of others and to enlarge the body of shared and agreed-upon understanding.

The first issue on which Dr Erwin and I disagree is that concerning science and scientism. Its bibliography is enormous, and Dr Erwin is as capable as I am of ferreting it out. I shall mention here only two classic books: Hayek's *The Counter-Revolution of Science*; and *Scientism and Values*, edited by Helmut Schoeck and James W. Wiggins. The second general area of our disagreement concerns the attempts which have been made for years to establish the genetic and/or organic base for schizophrenia. From the perspective of standards of evidence operative in the hard sciences, I have not found these efforts, as described in the literature on the subject, to be convincing. To my mind, Thomas Szasz—who was trained as a physicist—makes a much stronger case for the opposite point of view, outlined in his *Schizophrenia* (1976), which I reviewed in this journal in June, 1979. I do not exclude the possibility that such efforts might eventually prove successful. If they did, however, schizophrenia, as Szasz has frequently argued, would then be a physical disease, on the order—say—of tertiary syphilis. On the question of whether I misunderstand Dr Erwin's argument or simply state it succinctly, I leave the reader to judge for himself, keeping in mind only that I had stipulated at the beginning of my review (pp. 85–6) that Dr Erwin's 'consistent posture is one of precarious balance, presenting arguments on both sides of nearly every issue he examines', and that 'statements which . . . sound forthright [in my review] are qualified by numerous restrictions in the original'. Dr Erwin's entire passage ([1979], p. 147) which I compressed and quoted on pp. 90–1 of my review reads as follows:

> Rosenthal (1971, p. 84) concludes that the current evidence is sufficiently strong that the issue of whether genetic factors contribute appreciably to the development of schizophrenia must now be considered closed. Whether or not the issue is closed, the new evidence when combined with the findings from the earlier (and inconclusive) studies of monozygotic twins (reviewed in Rosenthal, 1971) is impressive. Our best bet at present is that at least one mental disorder, schizophrenia, results largely from organic factors.
>
> Schizophrenia also possesses the other important characteristics referred to earlier. It has a relatively specific set of symptoms (Franks, 1975); it tends to be disabling; and the symptoms are involuntary. On these grounds, then, it does seem plausible to extend the notion *illness* to cover schizophrenia. If that is so, and because the main symptoms (at least for certain types of schizophrenia) are mentalistic, then the case is strong for concluding that there is at least one mental illness.

With regard to his discussion of free will in Section §6, I did not claim that Erwin (1982, p. 57) was 'trying to prove that soft determinism is true'. What I said (p. 92) was that he solved the problem 'by opting for ***soft determinism***, which allows free choice'. On p. 57 of his Discussion, he says he was 'trying to prove [just] that soft determinism . . . is a plausible option that would have to be ruled out'. I fail to see where 'opting' is not a fair paraphrase for 'trying to prove . . . a plausible option'. (Actually, he had hypothecated ([1979], p. 178) soft determinism as 'a viable option'.)

Finally, in §10, Dr Erwin criticizes me for correcting the publication date of his book, implying that my 'source' was incorrect. Here, I am afraid, Dr Erwin is simply ignorant of technical bibliographical matters. My 'source' was a flyer from Cambridge University Press indicating that the 'publication date' was 8 March 1979. If I must choose between a flat statement by the author and a flat statement by the publisher as to a book's publication date, I must opt for the authority of the publisher. The reason I 'wanted to make an issue of the matter' once again involved a compliment I was paying Dr Erwin which he failed to recognize as such. Should his book drift into oblivion—and of course it may—then I would agree that the publication date is a mere quiddity. But it seemed to me that it may well become a minor classic in its field, and that future scholars, for reasons neither Dr Erwin nor I can anticipate, might find the actual publication date of some significance. For this reason I wanted to get that date on the record.

Presumably the editors of *The British Journal for the Philosophy of Science* are fully competent to defend themselves against the *ad hominem* attacks Dr Erwin levels against them. I will say only that the angle of vision from which he writes his 'Discussion' reflects the same naïveté about editorial procedures as it does about much else. Being uninformed is not a cardinal sin, but it *is* culpable to use one's ignorance as a platform for attacks on fellow scholars. The house of learning has many mansions, and Dr Erwin and I inhabit different ones. But surely we can walk down a corridor together without biting and scratching each other. I did (and do) consider his book to be, with some reservations, quite good of its kind. And I make that point clear, both at the beginning and the end of my review (pp. 85, 94): 'Most clinical psychologists who call themselves ***behaviour therapists***, or who are sympathetic to many or at least to a significant number of the cluster of techniques professionally recognized as characteristic of behaviour therapy, will want to read this book'. It 'warrants close attention in the study . . . by clinical psychologists and interested psychiatrists', being 'probably the most thorough book on the subject that is apt to be available for years to come', and it 'will undoubtedly have bibliographical significance'.

Whatever justification Dr Erwin presents for his ill-considered reply to my review, I cannot escape the conviction that it was triggered by the fact that not only did I understand both the issues and his book; I understood them in a broad enough context to point out some of their long-range implications. In any event, and regardless of the well-springs of his motivation, conscious or unconscious, I stand by my review. It is a detailed account of his book, in which I discuss his style, his methodology, outline at some length his substantive findings, and assess not only the weaknesses but also the strengths of his arguments. The irony is that in the long run, such a review will probably do his book more good than the hosannahs of unalloyed praise he seems to expect as his due.

M. E. GRENANDER
The Institute for Humanistic Studies,
State University of New York at Albany

REFERENCES

[1969]: *A Manual of Style*. The University of Chicago Press.
BAIER, K. [1981]: 'The Ethics of Behavior Modification', *Ethics*, **91**, pp. 499–509.
COMMISSION ON THE HUMANITIES [1980]: *The Humanities in American Life*. University of California Press.
ERWIN, E. [1979]: *Behavior Therapy: Scientific, Philosophical, and Moral Foundations*. Cambridge University Press.
ERWIN, E. [1982]: 'Behaviour Therapy: A Reply to an Incautious Discussion', *The British Journal for the Philosophy of Science*, **33**, pp. 53–60.
GOFFMAN, E. [1962]: *Asylums: Essays on the Social Situation of Mental Patients and Other Inmates*. Aldine Publishing Co.
GOULD, P. [1981]: 'So Human a Science', *The Sciences*, **21**, pp. 6–9, 28, 30.
GRENANDER, M. E. [1979]: 'Psychosis: Disease or Rhetoric?' *The British Journal for the Philosophy of Science*, **30**, pp. 177–82.
GRENANDER, M. E. [1981]: 'A Cautious Overview of Behaviour Therapy', *The British Journal for the Philosophy of Science*, **32**, pp. 85–94.
HAYEK, F. A. [1955]: *The Counter-Revolution of Science: Studies on the Abuse of Reason*. Free Press of Glencoe.
HILL, G. B., (*ed.*) [1934]: *Boswell's Life of Johnson*, revised L. F. Powell. 6 vols. The Clarendon Press, **1**, p. 471.
SCHOECK, H. and WIGGINS, J. W. (*eds.*) [1960]: *Scientism and Values*. D. Van Nostrand Co.
SPERRY, R. W. [1982]: 'Some Effects of Disconnecting the Cerebral Hemispheres', *Science*, **217**, pp. 1223–6.
SZASZ, T. [1976]: *Schizophrenia: The Sacred Symbol of Psychiatry*. Basic Books, Inc.

THE PHILOSOPHICAL RELEVANCE OF 'TECHNICALLY GOOD' EXPERIMENTS

Franklin [1981], in his paper "What makes a 'good' experiment?" discusses a number of ways in which an experiment can be 'good'. He also distinguishes experiments which are 'conceptually important' from experiments which are 'technically good' although some experiments can be both. It is a pity that Franklin dwells too briefly on technically good experiments. He gives examples and points out that they are those experiments 'in which improvements in existing apparatus or the construction of an entirely new apparatus result in a more accurate and precise measurement of a physical quantity'. Perhaps a more detailed discussion would have revealed that between the two kinds of technically good experiments—those that depend on 'improvement in existing apparatus' and those that involve 'the construction of an entirely new apparatus'—there can be very significant difference. While it is obvious that to be accounted good they both have to produce more accurate and precise measurements than hitherto, the second kind is also good in a very special sense. This is because before new apparatus is 'constructed', they have to be *designed*. And the second kind of technically good experiments can be good in that the improvement in the accuracy and precision of their measurements is the result of *ingenuity* in design. In other words, the improvement in accuracy and precision would not have come about if the authors of these experiments had not struck upon the 'bright ideas' behind their designs. This kind of experiment represents a kind of technical breakthrough, a quantum jump in the development of know-how in experimental science, and is therefore 'good' in that sense.

Technically good experiments of the second kind have been too little noticed by philosophers. Somebody who reads the philosophy of science who does not have a background in the sciences could very easily derive the impression that the carrying out of experiments by scientists is a routine matter possessing no glamour except when their results clash with some existing theory. But in fact, of course, experimental science is much more exciting than such a layman would think. Philosophers of science would therefore do a great service to their public if they remind them from time to time that ingenuity is required not only in the 'invention' of theories but also in *designing* experiments.

This aspect of science it seems to me has been too little looked into by philosophers, *that experiments have to be designed*. By definition, there can be no naturally occurring experiments. In any case, *technical* ingenuity is often required *even* in making *observations*. Experiments are questions we pose to nature. But nature does not speak our 'natural' or scientific languages. It is therefore not always easy to get through to nature with our questions. To have succeeded in talking to nature can be as exciting and exhilarating as to have succeeded in talking to somebody with whom you do not initially share any common language. This is why we value the answers we get back, inarticulate though they be.

The fact that experiments have to be designed has an important bearing on the Duhem-Quine thesis and this I want to bring out at this time. The Duhem-Quine thesis claims that theories are not refuted singly but in groups. For example, when an experiment is constructed to test a theory, other theories (called auxiliary theories) besides the theory under test are involved in deriving the observational statement, the 'prediction', to be compared with the actual result of the experiment. When the 'prediction' is found to be false, we cannot say, according to Duhem and Quine, that the theory under test is false but only the conjunction of all the theories involved including the theory under test. This thesis is unassailable when we confine our attention *only* to the manner in which the 'prediction' is derived.

However, we have to remember that before the 'prediction' can be derived, the experiment has to be designed. Now, if all the auxiliary theories on which an experiment depends are always as uncertain or questionable as the theory to be tested, then there simply cannot be such a thing as 'designing an experiment'. An architect in designing a building cannot be totally in the dark concerning the strengths of the materials he is going to employ. If he were, he would not be able even to start! Thus the practice of *designing* experiments implies that there is a certain amount of knowledge concerning the relative reliability of theories that function as auxiliary theories in experiments. It would make no sense to design an experiment based purposely on dubious auxiliary theories. As much as he can, a scientist designing an experiment will try to base it on theories that he thinks, based on past experience, he can rely upon.

We can compare the situation here with circumstances in which theories are 'invented'. Theories are invented to answer problems. Problems have to be formulated. Now if we do succeed in formulating problems at all, the statements in terms of which the problems are formulated cannot all be as problematic as the problems we intend to formulate. In other words, *before* a problem can be formulated, some estimate has to be made concerning the relative problematicity of various elements of our background knowledge; we would base our formulation of the problem as much as we can on that part of our background knowledge which is the least problematic. Now all our background knowledge is *in principle* revisable but if they are in our judgment all *equally* problematic then it will be quite impossible to formulate our problem and we would have no idea what theory to invent. To use a simple example, *before* we can ask the question, who is the murderer? we would like to be *relatively* certain that a murder has been committed. And although it is true that we are sometimes mistaken in thinking that a murder has been committed, we cannot on that basis regard our judgment as to whether a murder has been committed as *always* unreliable. Sometimes we are rather sure a murder has been committed from the evidence we have been able to gather, at other times we may be less certain because of lack of evidence or because of its ambiguity. When we are sure a murder has been committed, we then have a 'definite' problem on our hands: who is the murderer? When we are not sure, our problem is less definite.[1] Similarly, there are 'good' experiments for testing a theory and experiments which are less good.

A 'good' experiment for the purpose of testing a theory is one in which the auxiliary theories being used are relatively reliable, so that when the result of the experiment contradicts the 'prediction', we can say with a fair bit of confidence that the theory under test is at fault. Because designing experiments is not always an easy matter, it often happens that the best test we can come up with is an experiment which contains among its auxiliary theories *one or a few* whose reliability is relatively less certain. Thus it sometimes happens that the falsity of the 'prediction' is found to be due not to the theory under test but to some of the auxiliary theories. Hence a theory in danger of being refuted can be 'rescued' *sometimes*—but only sometimes—by appeal to the Duhem-Quine thesis. However, experimental science would not have been possible if the Duhem-Quine thesis were applicable in *every* case and can be appealed to whenever one wishes. It is worth repeating that a new experimental design adds to our experimental know-how: it could be further developed and exploited and could suggest *new ideas* as to how experiments for other purposes could be carried out. (Perhaps the cathode-ray tube is a good example here. How many experiments have made use of it as their centrepiece, or of apparatus of a similar design?) Inasmuch as an indiscriminate appeal to the Duhem-Quine thesis would make nonsense of the practice of *designing* experiments, such an appeal would not be compatible with an interest in the *advancement* of experimental science. It is therefore not a matter of convenience which theory we want to revise when there is a refutation.

---

[1] One does not necessarily have to have a perfectly definite problem first before going on to look for a solution. One may not be very sure that there has been a murder but may carry out an investigation nevertheless, on the assumption that there has been. One may even then find a suspect who *confesses* that he is the murderer, providing at the same time corroborating evidence!

Technically good experiments are good reminders of the fact that experiments have to be designed. They should be of interest not only to scientists but also to philosophers.

I should not conclude without adding one further word. I do not wish to imply by the foregoing that well-entrenched, usually-regarded-as-reliable auxiliary theories are not *sometimes* revised instead of theories under experimental test. They are; but such revisions would have been the result of investigations more stringent and extensive than usual, involving re-examination of theories other than those under the lime-light. Without such an extensive reappraisal, experimental scientists would find it difficult to practice their art. This kind of case finds its parallel in engineering practice. A roof collapses or a jetliner explodes in the sky. Faulty design is at first suspected but none can be found: none, that is, that can account for such a major failure. Suspicion may then arise that certain *current* engineering practices may not be as sound as they have appeared. Until these have been located, designers tend to be, or maybe are even by law required to be, very conservative in their designs.

From a logical point of view, a theory can be defended come what may. From a practical point of view, what theory to give up is not just a matter of convenience, but a question of moving ahead or remaining still.

TYRONE LAI
Memorial University of Newfoundland

REFERENCES

FRANKLIN, A. D. [1981]: "What makes a 'good' experiment?", *British Journal for the Philosophy of Science* 32, pp. 367–74.

Brit. J. Phil. Sci. 35 (1984), 161–173 *Printed in Great Britain*

Review Articles

A BAYESIAN APPROACH IN THE PHILOSOPHY OF INFERENCE*

Philosophical disputes that are not dissolved by Joad's law, 'It depends what you mean', often yield to Good's law, 'It's a matter of degree'. But Gilbert Ryle, one-time ruler of British philosophy, in a letter in 1947, said 'degrees of belief' are meaningless because you either believe something or you don't. In that letter he did not suggest an alternative phrase to capture the concept, such as 'intensity of conviction' or 'personal probability' or 'subjective probability'. If Ryle were right in a *substantive* sense there would be no intermediate stage between belief and disbelief and a change from one state to the other could never occur gradually. Common law seems to agree with Ryle because an accused has to be 'found' guilty or innocent if a decision is reached at all, but, as Mr Bumble in Oliver Twist said '. . . the law is a ass, a idiot'. If a jury reached a unanimous verdict that the probability of guilt was between 0.8 and 0.95, it might be charged with contempt of court.

Many statisticians during the last few decades have decided that degrees of belief are necessary for clarifying the foundations of inference, and in the last ten years some professional philosophers have come round to this view. Horwich is one of them regarding various puzzles and paradoxes in the philosophy of science.

The formal use of degrees of belief is often described as 'Bayesian' because of a couple of obscurely written papers by the Rev Thomas Bayes, published posthumously in 1763, but as usual the eponymy is misleading because Laplace's later but still pioneering contributions were much greater.

There are many interpretations of 'Bayesian', but the essential ingredient is the acceptance of certain axioms for rational degrees of belief, or epistemic probabilities. These axioms have received several justifications, beginning with lectures of Ramsey (1926). Some of us regard interval-valued probabilities (upper and lower probabilities) as more fundamental than sharp numerical ones, but, for the sake of simplicity, numerical values are usually assumed and will be assumed in this review. The current modern notation for a probability, in works on statistics and mathematical probability, is $P(A \mid B)$, read 'the probability of A given B or relative to B', or the 'conditional probability of A given B'. In practice all probabilities are conditional, but B is often taken for granted and then the notation is abbreviated, with some risk of error, to $P(A)$. Many philosophers of statistics use other notations because they don't read statistical texts.

The main usual axioms of probability, whether they are regarded as epistemic or physical, are (i) $0 \leqslant P(A) \mid B) \leqslant 1$, (ii) the additive law $P(E$ or $F) = P(E) + P(F)$ if $(E \,\&\, F)$ is impossible, and (iii) the product law, $P(E \,\&\, F) = P(F) \cdot P(F \mid E)$.

---

* Review of Paul Horwich [1982]: *Probability and Evidence*. Cambridge University Press.

The book by Horwich is the third book known to me with a similar title, the others being Ayer (1972) and Good (1950). The present book is closer to the latter than to the former work in its basic philosophy. The basic philosophy is also close to that of Jeffreys (1937, 1939) who in this century did as much as anyone to reinstate Bayesianism for the foundations of statistics and philosophy of science, though these foundations are still controversial. But Jeffreys and Laplace are not mentioned by Horwich.

At the Second World Congress of the Econometric Society in 1970, Jeffreys remarked that it had taken fifty years for his Bayesian views on the foundations of inference to have much impact among statisticians, and he predicted that it would take another fifty years before they had a similar impact among (professional) philosophers. The publication of Horwich's book and of books by Richard Jeffrey, Isaac Levi, Roger Rosenkrantz, and Teddy Seidenfeld are steps in this direction, but Horwich's book is addressed primarily to professional philosophers, and hardly at all to statisticians.

The first sentence of the book states that the book is about scientific knowledge, especially the concept of evidence. In view of this opening, and of the title of the book, let's first consider how 'weight of evidence' ought to be defined. Suppose then that H is a hypothesis or theory, for example, that an accused person is guilty, and let E denote the evidence presented in court. Let $W(H:E \mid G)$ denote the weight of evidence *concerning H provided* by E *given* G all along. For example, G might include what is known about guns in general, and the fact that firing guns at people is dangerous to their health. If we take G for granted we can abbreviate the notation to $W(H:E)$. How should it be defined in terms of probability?

It can be shown, under compelling assumptions, that $W(H:E)$ must depend only on the ratio $P(E \mid H) \div P(E \mid -H)$, where $-H$ denotes the negation of H, that is, that the accused is innocent (Good, [1968], pp. 127 and 141). This ratio is equal to the factor by which the initial odds of H are multiplied to obtain the final odds and is known as the *Bayes factor* in favour of H provided by E. (The odds corresponding to a probability p are defined as $p \div (1-p)$.) The Bayes factor was basic in Jeffrey's work on statistics, though he gave it no name. By taking $W(H: E)$ as the logarithm of the Bayes factor we obtain the intuitively satisfying formula

$$W[H:(E \,\&\, F)] = W(H:E) + W(H:F \mid E), \tag{1}$$

or, in words, the *weight of evidence concerning H provided by E & F combined is equal to the weight provided by E plus that provided by F when E is already allowed for*. We can therefore think of weights of evidence as weights in the scales of the Goddess of Justice, positive in one scale and negative in the other.

The concept of weight of evidence, though not under that name, was basic in Turing's approach to the cryptanalysis of the Enigma, in Mosteller and Wallace [1964], and in Tribus [1969]. I predict that in fifty years from now, if civilization is restored, $W(H:E)$ will be accepted by the majority of philosophers as by far the best explication of weight of evidence, and that after a further fifty years it will have a large influence in elementary education, in medical differential diagnosis, and in common law.

It is unfortunate that Horwich ignores this 'explicatum' of weight of evidence or confirmation. On page 53 he suggests the explicata $P(H|E)-P(H)$ and $P(H|E) \div P(H)$, the latter of which he mentions had been used by J. L. Mackie. It is equal to $P(E|H) \div P(E)$. Neither proposal is a function of the Bayes factor. Mackie's measure is what Keynes [1921] called the coefficient of dependence or association factor between E and H. That it is inappropriate as a measure of corroboration, confirmation, or weight of evidence, whether or not logarithms are taken, is readily seen. For if the probability (instead of the odds) of guilt is doubled, from 0.001 to 0.002, we would not be impressed, but if it were doubled from 0.499 to 0.998 we would be; and if it were doubled from 0.8 to 1.6 we would know we had gone mad. A similar objection applies to $P(H|E)-P(H)$. Although the logarithm of the association factor also satisfies equation (1), and has important uses in communication theory, and although $P(H|E)-P(H)$ also satisfies this equation, neither expression captures the notion of weight of evidence or corroboration.

When attacking Mackie's views the author does not notice that Mackie's formula for degree of corroboration or confirmation is a cause of much of the trouble (page 58).

Another topic given prominence in the book is Hempel's paradox of confirmation, or the raven paradox, so let's discuss it next. Consider the hypothesis H that all ravens are black, where a raven has been defined without reference to its colour. Let us call the sight of a black raven a 'case' of this hypothesis. This might appear to corroborate H. But equivalent to H is the hypothesis that all non-black things are non-ravens. Therefore the sight of a white shoe also corroborates H, and this seems absurd. The paradox was resolved by Good ([1961], [1967*b*]) and, in an ingenious and novel manner, by Horwich [1978], his argument being presented also on pages 54 and 55 of the book. It is proved in these places that if an object is selected 'at random' and is found to be a black raven this does *not* necessarily corroborate H: it depends on the background assumptions. The reason it seems to do so is that if we first sample a raven at random, and *then* see it is black, this does corroborate the hypothesis. On the other hand, if a black thing (B) is sampled at random and is found to be a raven (event R) this can provide a small positive or negative weight of evidence in favour of the hypothesis (H) that all ravens are black, depending on the background knowledge. The author is mistaken when he states on page 58 that the evidence in this case is strictly zero. Note that even $W(H:R)$ (and $W(H:B)$) need not be zero because, if a *thing* is sampled at random and is found to be a raven (or black) we obtain some evidence that ravens (or black things) are numerous. The background knowledge might show that this abundance gives evidence for or against H.

The concept of surprise is another topic given some emphasis. Horwich rightly says that an event E is surprising if (*i*) $P(E \mid C)$ is small, where C denotes our previous beliefs, and at the same time 'there is some initially implausible (but not wildly improbable) alternative view K about the circumstances, relative to which E would be highly probable'. This may be compared with the remark by the reviewer '. . . we tend to be surprised . . . when the result of an observation [E] has much greater probability when given some other, not entirely untenable, hypothesis [K instead of C]' (Good [1956], p. 1131). But I do not blame Horwich for overlooking this remark, made in a statistical periodical, for there is only one person who has read everything I've written!

Horwich devotes a chapter to the tricky problem of scientific induction. If n objects of kind A, sampled at random, have been found to be B's, then the next one, sampled in a similar manner, will probably be a B, and the probability increases to 1 as n tends to infinity. Let's call this the principle of *predictive* induction. Stated for *all* future A's, instead of just for the next one, the principle may be called that of *universal* induction, a principle that is less important and less reliable than that of predictive induction. Russell ([1912], p. 69) says 'If the principle [of predictive induction] is [totally] unsound, we have no reason to expect the sun to rise tomorrow . . . or to expect that if we throw ourselves off the roof we shall fall'. To which we could add 'or to understand the denial of the principle' *for we use induction when we assume that words mean what they have meant in the past* (compare Black [1967], pp. 178–9; and Good [1981]).

Goodman [1955] propounded a 'new riddle of induction': if all emeralds (defined without reference to colour) have been green this might seem to support the hypothesis that all emeralds are green; but it equally supports the hypothesis that they will be 'grue', that is, green until the year 2000 and blue thereafter. In fact the weight of evidence for distinguishing these hypotheses is zero. The Bayesian resolution of this 'gruesome' paradox, and there are no other resolutions, is of course that the *prior* probability of gruesomeness is extremely small, as pointed out, for example, by Horwich and also by Good ([1975], pp. 55–56) where earlier references are given.

In his discussion of scientific induction Horwich refers to some ideas of Russell [1912], but Russell ([1948], p. 426) says the best mathematical treatment is that of Keynes ([1921], p. 235) which is essentially the same as that of Wrinch and Jeffreys [1921] or Jeffreys ([1939], p. 38). Jeffreys's conclusion was that 'repeated verification of consequences of a hypothesis will make it practically certain that the next consequence of it will be verified'. But this would not be true if the new circumstances were known to be of a different character from the previous ones (for example, black swans were found in Australia), a point not mentioned by Jeffreys, but emphasized by Keynes ([1921], pp. 219–20). In relation to this point Horwich (p. 121) says '. . . evidence for a hypothesis is . . . significantly diverse to the extent that it is improbable, relative to those alternative hypotheses with high prior probabilities'. In fact it is difficult to establish high probability for a hypothesis H, based on a sequence of observations $E_1, E_2, E_3, \ldots$, unless one is able to state other hypotheses clearly enough to be able to judge $P(E_{n+1} \mid E_1 \,\&\, E_2 \,\&\, \ldots \,\&\, E_n \,\&\, -H)$ (Good [1975], p. 62). Indeed the concept of weight of evidence subsumes that of induction.

On pages 122–9, the problem is discussed of why extra observations are of expected value in the evaluation of a hypothesis, if the cost of acquiring the observations can be ignored. Horwich cites Good [1967*a*] where the answer was based on the economist's notion of expected utility, and Horwich then gives a 'purely epistemological solution'. He defines '*the error in a probability evaluation* as the difference between that value [p] and the truth value (1 or 0) of the hypothesis in question'. Thus he really introduces a negative utility or cost but it depends on the asserted probabilities, and can be called an epistemic utility, whereas the costs in Good's account measured the losses through performing various acts given various hypotheses. Horwich proves, by a page of algebra, that additional evidence is always of expected value (in the sense of mathematical expectation). Actually a more general result can be proved. Note first that if the epistemic utility or value of asserting p, when an event E is true, is some function of p, say $f(p)$, then for consistency it must be $f(1-p)$ when $-E$ is true *i.e.* when E is false. It can be shown that additional evidence is then always of expected non-negative value if and only if the graph of $xf(x)+(1-x)f(1-x)$ is convex when viewed from below. We get Horwich's case when $f(x) = x$. A generalization to more than two hypotheses is readily made.

The last chapter contains a defence of inductivism against Popper's attack on it. As Horwich says, Popper failed to recognize 'that content and probability may each be virtues of a scientific theory, even though it is difficult to optimize both of them at once'. A compromise is necessary and has been proposed, as a sharpened form of Ockham's razor, by Good ([1968], [1975], [1977]), where applications to statistics are made. The topic might one day be transferred from the departments of philosophy to those of statistics (compare Russell [1912], p. 155).

In summary, this is an intelligently written essay on the power of the Bayesian approach in the philosophy of inference, but it ignores too much good work.

I. J. GOOD
Virginia Polytechnic Institute and State University

REFERENCES

AYER, A. J. [1972]: *Probability and Evidence*. Columbia University Press.
BLACK, Max [1967]: 'Induction'. In *The Encyclopedia of Philosophy*, 4, 169–81.

Good, I. J. [1950]: *Probability and the Weighing of Evidence*. Charles Griffin; Hafners.

Good, I. J. [1956]: 'The surprise index for the multivariate normal distribution', *Annals of Mathematical Statistics*, **27**, 1130–5.

Good, I. J. [1961]: 'The paradox of confirmation, II', *British Journal for the Philosophy of Science*, **12**, 63–4.

Good, I. J. [1967*a*]: 'On the principle of total evidence', *British Journal for the Philosophy of Science*, **17**, 319–21.

Good, I. J. [1967*b*]: 'The white shoe is a red herring', *British Journal for the Philosophy of Science*, **17**, 322.

Good, I. J. [1968]: 'Corroboration, explanation, evolving probability, simplicity, and a sharpened razor', *British Journal for the Philosophy of Science*, **19**, 123–43.

Good, I. J. [1975]: 'Explicativity, corroboration, and the relative odds of hypotheses', *Synthese*, **30**, 39–73.

Good, I. J. [1977]: 'Explicativity: a mathematical theory of explanation with statistical applications', *Proceedings of the Royal Society (London) A 354*, 303–30.

Good, I. J. [1981]: 'Can scientific induction be meaningfully questioned?', C100 in *J. Statist. Comput. & Simula*, **13**, 154.

Goodman, Nelson [1955]: *Fact, Fiction and Forecast*, third edn. Bobbs Merrill.

Horwich, P. [1978]: 'A peculiar consequence of Nicod's criterion', *British Journal for the Philosophy of Science*, **29**, 262–3.

Jeffreys, Harold [1937]: *Scientific Inference*. Cambridge University Press (2nd edn., 1957).

Jeffreys, Harold [1939]: *Theory of Probability*. Oxford University Press (3rd edn., 1961).

Keynes, J. M. [1921]: *A Treatise on Probability*. Macmillan.

Mosteller, Frederick and Wallace, David L. [1964]: *Inference and Disputed Authorship: The Federalist*. Addison-Wesley.

Ramsey, F. P. [1926]: 'Truth and probability' reprinted in *Studies in Subjective Probability* (H. E. Kyburg and H. E. Smokler, eds.; Robert E. Krieger, 2nd edn., 1980), 25–52.

Russell, Bertrand [1912]: *The Problems of Philosophy*. Oxford University Press.

Russell, Bertrand [1948]: *Human Knowledge, its Scope and Limitations*. George Allen and Unwin.

Tribus, Myron [1969]: *Rational Descriptions, Decisions and Design*. Pergamon Press.

Wrinch, Dorothy and Jeffreys, Harold [1921]: 'On certain fundamental principles of scientific enquiry', *Philosophical Magazine, series 6*, **42**, 369–90.

T. S. KUHN, FUNCTIONALISM, AND SOCIOLOGY OF KNOWLEDGE[*]

The title of Barry Barnes's *T. S. Kuhn and Social Science* is too ambitious and far beyond its scope. Something like 'T. S. Kuhn and Sociology of Knowledge' might have been a better title, had there not been an extensive modification of Kuhn's own views in the argument. Perhaps, 'Kuhn and the functionalist approach to sociology of knowledge' would be a more accurate description of the book's content.

The claim that there had been no sociology of knowledge prior to Kuhn's historical generalisations about science and the scientific community (and the author's functionalist interpretation of them) is obviously incorrect. Marx, Engels, Lucacs, Manheim and so many others, who wrote and flourished long before Kuhn and functionalism, are thus ruled out of existence by definition. Indeed, it is 'the' sociology of knowledge to which the author regularly refers, and which he frequently defines in words such as the following:

---

[*] Review of B. Barnes [1982]: *T. S. Kuhn and Social Science*. Macmillan. £12.00 hardback. £3.95 paperback. Pp. 135.

> The sociology of knowledge is a matter-of-fact, empirical field of study which happens to include, among its subject-matter, the knowledge and culture of science (*Preface*, p. xi).

Take Manheim's ('the'?) sociology of knowledge, for example. This, too, pertains to matters of fact, although— rather like Barnes's different approach— its scope of application is apparently so wide that it looks more like a tautology than a theory. But the author's emphasis on the 'matter-of-fact' quality of his approach is meant to contrast with the normative rules of philosophy of science, and especially those of Popper.

Two points seem to have gone amiss, however: one, that Kuhn himself openly identifies the *criterion* of testability as a line of *demarcation* between science and non-science (although Barnes apparently rejects this, *cf.* p. 90); two, that while norms and criteria are free of time-spatial constraints and can be held *as such* in all times and places, 'matters of fact' must necessarily refer to given time-spatial contexts: they must have been changing throughout history, and they are likely to change in the future.

Barnes rejects universalism, but one of the weakest aspects of Kuhn's as well as his own views is their universalism in time and space: they say nothing about the specific circumstances in which their proclaimed 'matters of fact' about science and scientists (as well as non-science and non-scientists) would hold, as if these facts have been given to us from the Fall until Doomsday. Thus, the author's unexceptionable view that history should not be read backwards (which should also mean that we should not read our own views into the whole of intellectual history) could have been put to good use in his own exposition of the functionalist sociology of knowledge.

History is important both as a source of information about the past and as an indirect guide for an understanding of the present. Therefore, Barnes's emphasis on the value of historical knowledge ought to be welcomed. But his further claim that prior to Kuhn's theory this had not been realised by philosophers and social scientists, once again, flies in the face of the facts. Classics aside, one needs only to mention Machiavelli, Vico, Montesquieu, Herder, Michelet, Adam Smith, Richard Jones and Marx among many others to remind the reader that, not only a regard for historical knowledge, but— more significantly than that— conscious efforts for an *understanding* of historical facts (as opposed to backward readings of history) have been long with us.

Barnes's equal emphasis on the cultural dimensions of social phenomena— the variety of historical experience, in spite of its broader similarities— is also commendable, especially as it comes from a functionalist sociologist. But it runs counter to his own reduction of all previous approaches to history to that of the English Whig historians— a blatant piece of cultural reductionism which reminds one of G. B. Shaw's light-hearted metaphor of the barbarian who regards the prejudices of his tribe as laws of nature. There have been various approaches to history (many of them not even English or British) other than that of Whig historians, throughout history itself. Indeed, the approach which Barnes seems to regard as Kuhn's invention is (in the more recent times) akin to that of Ranke and his famous school, and well-known British examples of it are to be found in the works of Namier and others.

True, a regard for historical knowledge and understanding has often been derided by 'normal social scientists', among whom functionalist sociologists of the fifties and sixties tended to top the list. They saw 'facts' but ignored their own *a priori* preconceptions in selecting and interpreting them. And they paid little attention to history, culture and the social context, for example, in analysing 'voting behaviour' in America and Guatemala alike. Universalism in recent social science goes back to the conservatism of Positive Economics and End of Ideology; Popper, modern liberals and Marxists had very little to do with it.

But the author is aware, and he occasionally admits, that *facts-in-themselves* are not available to any of us; and that even the most explicit recognition of facts is bound up by the different *theories* with which we begin and end our discussions of them— in this case, *different* theories of sociology of knowledge (as opposed to 'the' sociology of knowledge). Thus he writes:

> Here as everywhere Kuhn writes as a functionalist and his account will only be acceptable as it stands to those who see no difficulties in accepting *the functionalist point of view* (p. 54, emphasis added).

It so happens that this reviewer has also put forward some propositions about sociology of 'normal science' with an even more open, explicit and comprehensive recognition of the facts (see Katouzian [1980], ch. 5). But, acceptable or not, they are clearly not based on 'the functionalist point of view'. Facts are facts, but theories are not. Roberto Michels's 'the iron law of oligarchy' is (at best) a theory for explaining the role and function of authority in social organisations—the state, its departments, political parties, trade unions, and perhaps even modern scientific establishments. But one does not have to ignore the (limited) factual content of his theory in order to reject its claim to being an 'iron law' of nature; or (unlike Michels himself) refuse to support *il Duce*.

Although the functionalist approach has been identified both with Kuhn's theory and with 'the sociology of knowledge' the author frequently uses conventionalist and instrumentalist arguments which are hardly new discoveries in sociology and philosophy of science. Conventionalists have tended to applaud the 'matter-of-fact functions' of scientific dogmatism (the author's own word) only a little less openly than does Barnes. Instrumentalists, on the other hand, place the emphasis on the functions (or 'usefulness') of given scientific dogmas— on 'whether they work or not'—

with equal enthusiasm. The question to which all of them fail to address themselves, however, is a comparison between the results—the function, utility, efficiency, etc.—of dogmatic (or normal) and undogmatic (or extraordinary) knowledge and science in history: between medieval schoolmen, for example, and the Renaissance men of art, science and intellect.

There is also a value judgement involved here. Every social activity and institution is based (and is bound to be based) on some discipline and authority (or—to use a more recent sociological jargon—on 'organisation and control'). But it does not follow that every existing discipline and authority give us a Panglossian 'best of all possible worlds'. Pedagogy (which heads one of the book's sections) is clearly important. Children do need to be brought up, but nearly all of them will be declared adults sooner or later. Education—both formal and informal—imply elements of authority as well as dogmatism, but this is a far cry from an unchanging dogmatism and authoritarianism from the cradle to the grave. And this should be least acceptable in the higher elevations of our self-proclaimed gropings for knowledge and science.

To the proclamation: 'that is how it is, like it or not', there is the answer: 'That is not how it has always been; it is not inevitably how it can be; and some of the best moments of scientific as well as social history have been spent in rejecting "what there is", and replacing it with something better.' Indeed, according to Kuhn's own theory of scientific revolutions, it is the great scientific achievements which afford the normal scientists their 'puzzle solving' activities.

However, Barnes has some interesting things to say about 'the functional necessity' of Kuhn's scientific revolutions which also say something about his own selective interest in Kuhn's theory:

> Kuhn, as we have seen, imbues the alteration of normal and revolutionary science with functional necessity (1970, ch. 9). Commentators accordingly describe his work as a 'theory of scientific development'. . . . This, however, even if it is a legitimate interpretation, is one which demands too much of Kuhn's account (p. 57).

This reviewer happens to be one of those commentators who are anonymously cited in the above quotation (see Katouzian [1980], ch. 4). Whether or not Kuhn himself has demanded too much of his own 'account', his text has unmistakably offered a theory of scientific progress, or a theory of the history of science. According to Kuhn, 'normal science' is characteristic of 'mature sciences', and *mature* sciences are those which are dominated by a single paradigm. If these specifications for paradigms (and paradigm change) in mature sciences now seem to be more-or-less true of the contemporary facts about *all* knowledge and science, then this shows up a weakness in the theory itself.

Another important feature of Kuhn's scientific revolutions is that they occur only as a result of an accute crisis as well as the emergence of an 'alternative paradigm' which is acceptable to 'the invisible college'. This would make scientific progress a rare phenomenon. Furthermore, since

Kuhn's concept of paradigm (and the factual examples which he gives) refer to piecemeal theories as well as epoch-making systems of thought, it follows that any change in any theory must come about through Kuhn-type crises, alternatives and revolutions. And this flies in the face of the facts even in periods which are marked by the high tides of normal science and scholasticism (see further, Katouzian [1982]).

Yet, Barnes's dislike of Kuhn's scientific revolutions does not exclusively serve an apologetic function for the 'anomalies' in Kuhn's own theory; it also stems from a conservative outlook towards scientific establishments which goes significantly beyond Kuhn's. In fact, he is much more interested in the 'functional necessity' of 'the structure of normal science', than that of 'the structure of scientific revolutions'. Perhaps that is a reason why he abbreviates the title of Kuhn's famous monograph to *structure* rather than *scientific revolutions*. It is 'normal science' which the author believes to be of fundamental importance, whereas, in Kuhn's monograph, 'normal science', 'crises', etc., are important ingredients in the emergence of scientific revolutions. The author's own preference for normal science may be more in line with his functionalist approach to the subject, but it tends to erase much that is distinctively due to Kuhn.

This reviewer has pointed out some of the basic weaknesses in Kuhn's scheme, elsewhere. Unfortunately— although there does seem to have been an awareness of the arguments— Barnes's book does not address itself to these questions. We call it unfortunate, because it makes it necessary to mention some of them briefly here. The most important problem is that of the meaning of the concept of 'paradigm' and its various 'functions' in Kuhn's theory. Barnes himself admits (though without further discussion) that, in this respect, 'difficulties remain'; and he simply chooses one of the various meanings given to it by Kuhn—that of an 'exemplar', 'an accepted problem-solution', 'a particular concrete scientific achievement'—throughout his book.

But, even if one wishes to follow a Kuhn*ian* (as opposed to Kuhn's own) approach, one would have to be able to defend the basic premises of the original theory. 'Paradigm' in Kuhn has very important 'functions': it is unique, it defies rational and empirical refutations even when these *actually* occur, etc., etc. Putting that aside and sticking to the notion of 'exemplar'—'the concrete scientific achievement'—there would still be a world of difference between Newton's mechanics as an 'exemplar' (or 'cluster of exemplars'), and one that relates to the production of a mechanical watch; between Darwin's 'concrete scientific achievement', and the numerous biological theories and experiments which have risen and fallen without much fuss before and after it. The persistence of such important ambiguities do not help the theory, and put in doubt its claim to general (let alone universal) applicability.[1]

---

[1] For a more extensive discussion of the above points, see further, Katouzian [1980], ch. 4.

To his credit—and in spite of his own lapses into it—the author disagrees with universalism both in and about science: 'Science is not a set of universal standards, sustaining true-descriptions of valid inferences in different specific cultural contexts' (p. 10). This non-universalist view of science is somewhat developed in the section on 'finitism' (pp. 35–40). The argument here is interesting, except when it implies that in order to reject universalism, one would have to be an instrumentalist in methodology, and a functionalist in sociology of science. Moreover, and in spite of his rejection of non-universalism, the author has completely ignored the universalist features of Kuhn's theory—features which threaten to turn it into a descriptive tautology about every 'normal' and 'extraordinary' situation under the blue sky, as opposed to a general theory of scientific 'normality' and revolution.

Barnes is much fond of using the label 'Manichaean' to deride the alternative views of others—the dualistic vision of the struggle between 'good' and 'evil' which Manichaeans had themselves borrowed from Zoroastrians, and which they partially passed on to other Middle Eastern religions. But he overlooks a similar dualism (or 'Manichaeanism') in Kuhn's vision of ('the good'?) normal and ('the evil'?) revolutionary science. The more important problem, however, is the apparent absence of a time-spatial context in this specific process of continuity and change. Furthermore, it looks like a description of normality and revolution anywhere and everywhere, in and out of science: of republic and revolution in Rome, and of Newton and Einstein in the twentieth century. Normal science in our time may have a function; and a description of its functions, too, may have another function. But the whole scheme as it stands looks like a universal tautology.

The author's emphasis on his 'matter-of-fact' approach begins to be abandoned when he comes to deal with the issue of 'moralising':

> Everybody moralises: there are no exceptions. But if moralising is an unavoidable biological weakness, conflating it with description is an outright crime—a crime which, knowingly or unknowingly, the Popperians frequently commit. Consequently, it is all too easy to mistake the impossible (and undesirable) prescription of Popperian epistemology for a well-informed exposition of the features of actual science (p. 60).

The contradiction between the opening and the closing sentences of this statement is the fundamental contradiction which runs through the whole of Barnes's argument. Few would doubt that scientific or social institutions have certain functions, and that these functions should be studied. But even if our descriptions of these functions were faultless, would we not be 'conflating descriptions with moralising' if we then concluded that what is, is right; that that is how it has been in science as well as society in all times and places; and that that is how it will be, and how it ought to be, in the future as well?

We cannot speak for Popperians for—to his credit—Popper has not offered a closed paradigm or a universalist tautology. As for Popper's own views, he suggested the *criterion* of falsifiab*ility* (which Kuhn has accepted); and *described* the process of falsific*ation* as a characteristic of 'extraordinary' science, which he further recommends *as a norm* to the contemporary 'normal science'. However, as someone who has explicitly recognised the facts in an 'exposition of the features of actual science', this reviewer has an equal right to reject the moral virtues that Barnes and others have claimed for them. 'A well-informed exposition of the features of actual science' does not have to give rise to the author's *unique* mode of 'moralising'. It can surely allow for an antithesis to such moralising as well.

Barnes speaks of the activities of normal science being 'intelligible as a pattern of conventional activity supported by authority and mechanisms of social control' (p. 65). But there can be various 'patterns', various types of 'authority', and various ways of applying 'social control'. To what purpose, in whose interest and how efficiently are these 'facts' working? One can equally describe social activity in the above terms. But is there no further distinction between democracy, aristocracy, dictatorship, or totalitarianism? Should a 'well-informed' study of a totalitarian state together with its 'patterns', 'authorities', 'mechanisms', etc.,— conclude that that is 'the best of all possible worlds' by virtue of the fact that it 'exists' and 'works'?

Besides, a 'well-informed exposition' of any organisation would surely compare its functions and results with its constitution— its self-proclaimed aims and objectives— and it would focus its attention as much on the position, views and welfare of the ordinary members of that organisation as on its ruling authorities. Barnes's argument is that 'normal science' exists and holds together because of the 'dogmatism' (his word), the 'authority', and the 'social control' which that authority uses in laying down the law. The point, however, is not about the desirability or undesirability of law itself, but about those who do and do not partake in laying it down; not about 'intolerable individuality' as opposed to order, but about the order which is needed to protect freedom of thought. If normal science is a reality, then so are the countless scientists who are its subjects. Should they and their views and interests not equally feature in a 'well-informed exposition' of the subject?

The trouble with the dogmatism which Barnes openly praises is that it shuts its eyes and ears to the weaknesses of its own dogma even when these are pointed out by others. This was the function of the old theologies which Renaissance science and philosophy managed to defeat. There is a world of difference between 'commitment' and blind obedience. Reason and evidence may be weak before power and authority, but this does not prove the scientific superiority of the latter. A genuine dialectic of facts and ideas is all that we basically possess for sorting out our differences, and groping towards an understanding of reality. This is closed to dogmatism, for it refuses to reorganise arguments from outside its own 'framework'. The clash of different dogmatisms may in fact end up by the triumph of our dogma against the others. But there will be no guarantee that this is the best dogma, or— even if it was— the best possible outcome for science or society. We need to face awkward questions, and, if need be, change our minds.

Barnes's functionalist sociology of knowledge should not be laughed out of court whatever our own contrary views, for this is the practice of dogmatic critics and reviewers. On the contrary— and in spite of the 'intolerable individualism' which purely self-interested attitudes often provoke— it should be read and discussed even more acutely by those who may find themselves in disagreement with it. It is an unsuccessful rescue attempt for a theory of scientific knowledge which is ridden with unresolved ambiguities and unbroken circularities. And it further exposes these problems in its more open exposition of the conservative ideology on which it is based. The fact that this has been done with intelligence as well as cleverness goes without saying.

HOMA KATOUZIAN
University of Kent at Canterbury

REFERENCES

KATOUZIAN, HOMA [1980]: *Ideology and Method in Economics*. Macmillan and New York: University Press.

KATOUZIAN, HOMA [1982]: 'The Hallmarks of Science and Scholasticism', in Nobert Elias, *et al.* (*eds*), *Scientific Establishments and Hierarchies*. Sociology of the Sciences, Yearbook 1982. D. Reidet.

Reviews

KYBURG, HENRY E., Jr [1983]: *Epistemology and Inference*. University of Minnesota Press. Pp. xi+318.

Here are nineteen of Henry Kyburg's papers, on probability and epistemology. Not all such collections are worth having, but this one is. Kyburg is single-minded and thorough, a specialist in a big and lively subject whose progress is well reflected in his work. His prose is moreover clean and pithy, pleasing as well as profitable to read. If he symbolises somewhat beyond necessity, it is only because symbols come naturally to him (as in his work they need to), not to show off. And if I and others dispute his conclusions, no one disputes the advancement of the subject by his arguments.

The core of Kyburg's work is probability, for him a logical and epistemological notion. He rejects the subjectivism often derived from decision theory, insisting that acquiring knowledge is not just an action made more or less rational by its expected utility. Rather, to be rational an agent first needs knowledge, including knowledge of the probabilities used to compute his expected utilities. But for Kyburg knowing a probability is not knowing an empirical fact. It is knowing a logical relation between a statement and the set of statements that constitute the agent's body of knowledge, a relation that constrains the degree to which he should then believe that statement.

So far, so like Carnap, Hintikka and others. But no further. They only add to the subjectivists' probability axioms; Kyburg amends them. First, his probabilities are not single numbers but intervals. The probability of a coin landing heads will not be a number p but an interval (p, q). Probabilistic degrees of belief should then lie in that interval, though no degree of belief within it will be better than any other. (Kyburg denies in any case that we have precise probabilistic degrees of belief, arguing especially that so-called 'Dutch Book' arguments show neither that we do nor that we should.) This in turn makes expected utilities intervals and the maxim 'Maximise expected utility' useless when alternative actions' utility intervals overlap.

Next, not all statements have probabilities. Probability statements don't, being (like deducibility statements) meta-language statements about logical relations between statements in an object language in which they themselves do not occur. Nor are Kyburg's probabilities fixed by counterparts in the object language, *viz.* by statements of objective chance. Instead they are fixed directly by statistical (*i.e.* frequency) statements, as follows (p. 140):

> . . . we will say that the probability of a statement S is the interval (p, q), when we *know*, in our general body of knowledge G, (1) that S is equivalent to a statenent of the form $a \in C$; (2) that a is a random member of B with respect to C; and (3) that the

proportion of B's that are C's lies in the interval (p, q). Furthermore, to know that a is a random member of B with respect to C, is to have in our general body of knowledge G the statement '$a \in B$', but not to have simultaneously in G any triplet of statements '$a \in B*$', '$B*$ is included in B' and 'the proportion of $B*$'s that are C's does not lie in the interval (p, q)'.

Getting 'single case' probabilities (of a being in the class C) from frequencies like this seems an advance on frequency theories. But arbitrary $B*$'s that correlate more or less with c's than B's do must still be ruled out. As Kyburg says (p. 210): 'The frequentist problem of choosing a reference class is . . . the same as the problem of defining randomness'. Either way probability is defined by relative frequency. Only whereas frequentists stick to real frequencies, Kyburg (p. 145) does not: '. . . the correctness of a probability statement demands only that the corresponding statistical statement be known— i.e. *reasonably believed*— rather than that it be true.'

According to Kyburg, laws and theories also lack probability. Instead he makes all accepted generalisations analytic (ch. 19), putting their empirical content into the error hypotheses used to dispose of counter-examples. The fewer errors the better, of course, so (unlike Kuhn and Feyerabend) Kyburg does offer an empirical way of assessing alternative generalisations. But why call them analytic? Kyburg says (p. 304) that 'to accept [a] generalisation *requires* interpreting the counter-examples as erroneous, and expunging them from our stock of moral certainties. And is not this the hallmark of analytic truths?' Well hardly, since it is the hallmark of all truths, and indeed of all members of Kyburg's bodies of knowledge, from which he rightly excludes pairs of contradictory statements (p. 245).

Kyburg, however, offers more than odd rewrites of old ideas. To him we owe the lottery paradox and its important consequence that we don't know, because we don't believe, the conjunction of everything we do know (ch. 14). To him (and Levi) we owe the point that probabilities— like other physical and psychological quantities— lack precise values, so that decision making needs more than the maximising of expected utility. To him we owe vivid proofs that different concepts of probability lead to different practical decisions; so that his resistance to perverse fashions, for subjectivism and for making epistemology part of decision theory, matters in practice as well as theory.

But Kyburg himself is not immune to perversity. There is his conviction (p. 123) that

> . . . what interests poker players as a basis for their probability judgments is the question of what proportion of one-card draws in the past or future history of poker will result in the completing of a flush. If there were an official 'chance' which differed from that frequency, I can see no reason why they should be interested in it.

As Kyburg says (p. 108), 'we have here a simple conflict of intuitions'. To me it seems perverse not to act on a known chance of completing a flush on this hand, whatever the known proportion of flushes in the whole Universe; just as it would be perverse to take only one box in the Newcomb situation

(Nozick [1969]), or to repent just to get into a class of people containing a higher proportion of those predestined to be saved.

Here, incidentally, Kyburg is hardly fair to subjectivists. A subjective probability *p* of completing a flush this time entails a very high subjective probability that the proportion of flushes in many similar cases will be about *p*. Kyburg calls this derivation of 'very strong beliefs about matters of fact on purely *a priori* grounds' (p. 68) 'evidence of *a priorism* of the worst sort' (p. 72). But why, when the conclusion is supposed to be as subjective as the premise and is clearly part of what it means? Subjective probabilities close to 1 or 0 are not especially absurd. The inference might indeed be run the other way (as Kyburg thinks poker players run it) from a conviction about how frequent flushes are throughout space and time to a single case subjective probability. But that conviction is hardly the one for which there is clear evidence that subjectivists can be accused of ignoring. It's the obvious objectivity of the single case probability that makes subjectivism incredible— *and* justifies convictions about frequencies of flushes in outer space.

Kyburg also treats Dutch book arguments for subjective probability rather perversely (ch. 6) by ignoring versions (*e.g.* mine!) that have anticipated and met his criticisms. But having argued the point in Part II of my [1980], I here simply repeat, *pace* Kyburg, that they do warrant a probability measure of the strength of beliefs; though Kyburg's offer of 'a probability concept which took "degrees" of rational belief to be measured by intervals, rather than by points' (p. 77) would in fact satisfy me.

Kyburg's basic objection to a betting measure of belief, however, is that though 'from the very outset designed as a theory of action . . . it is expressed in epistemic, or more properly credential, terms' (p. 20). This he thinks (ch. 2) reduces belief to behaviour, science to supplying merely instrumental aids to decision making, and epistemology to decision theory. But it doesn't. Measuring belief by how it affects betting behaviour is not behaviourism, just a modestly 'functional' characterisation of a state of mind. It doesn't make beliefs mere dispositions to action (Kyburg p. 22): since what my beliefs make me do depends on my desires (and *vice versa*), my actions cannot completely define either. Functionalism takes more than behaviour to characterise mental states: *e.g.* the perceptions that cause them (especially beliefs), and their own internal interactions. It has more resources than Kyburg allows, including some— *e.g.* the way a belief's standard perceptual causes can help to specify its content— that promise an epistemology altogether less ethereal than his.

Nor do betting measures of belief entail instrumentalism about scientific generalisations that can't be bet on. Betting needn't after all be the only measure of belief, any more than temperatures defined (say) by readings of ideal gas thermometers have to be measured that way. And anyway I can always ask myself at what odds I *would* bet on the truth of a statement (to be certified *e.g.* by God) if I could. Kyburg says that 'this puts us in a never-

never land of counterfactual psychological tests' (p. 207), but why is it a never-never land? All measures imply counterfactuals about what would happen if they were applied, and these need not be realisable to make sense or even to be known. Consider how known counterfactuals starting (*e.g.*) 'If there were a greatest prime number . . .' are used to prove their own unrealisability. Or how Newtonian gravitation makes his own First Law of Motion ('If a body were acted on by no forces . . .') unrealisable. Does this put mathematics and Newtonian mechanics in a never-never land? In psychology, even realisable betting counterfactuals are mostly known by introspection, not by actual betting; and nothing in functionalism prevents introspection informing me reliably of my own mental states (*cf.* Part I of my [1980]).

Functionalism can in short easily accommodate our conscious beliefs (and doubts) about scientific generalisations; *and* the desire to increase our knowledge of them for its own sake. That desire we can define as what makes us act to resolve our doubts, but only by means we believe likely to generate true beliefs (*e.g.* consulting telescopes rather than astrologers about the stars)—a proviso that distinguishes it functionally from a mere desire for more belief.

None of this commits us to a decision-theoretic epistemology. It does say that what we come to know depends partly on what we decide to find out about, but no one would deny that. It doesn't say we decide what to believe to what degree—or what to 'accept' into Kyburg's 'rational corpora', *i.e.* bodies of beliefs that rational men know are as probable (relative to their more certain corpora) as their concerns demand. We do neither of these things, not even in deliberation. Thinking out a conclusion is only a kind of action done to find something out: like looking to see something, but internal. And just as we can decide where to look for something but not what we see there, so though we can decide what to think about, we don't decide what conclusion, if any, we shall come to.

We may however still ask what beliefs our thought, and our senses, *should* deliver, to help us decide how to improve our ways of thinking and perceiving to deliver them. The ideal and uncontentious answer is of course that they should deliver *true* beliefs, truth being belief's whole object. But we cannot expect to get the truth all the time, so we try instead to ensure high probability, *i.e.* to instil habits of thought and perception that will make the beliefs they deliver very probable. But what that amounts to depends on how one construes probability; and here lies the real divide.

Kyburg restricts his epistemology to logical probabilities relating what we already know to prospective additions to our bodies of knowledge (*i.e.* of reasonable belief: Kyburg doesn't demand truth.) They are not enough. First, they can't cope with beliefs got by perception, whose probability depends on causal relations to what they are about, not on logical relations to existing beliefs. Second, they won't do even for new beliefs got from old ones by thinking. Obviously, thinking should make the beliefs it delivers probably true. But no logical relation to existing beliefs will do that unless they in turn are true, or at least probable in the contingent sense that Kyburg denies.

Kyburg's epistemology is of course, as he admits, incomplete. But he is committed to completing it without the contingent physical probabilities I and others think it needs. And this dispute about probability is no mere technicality. It embodies the most fundamental dispute in epistemology: *viz.* whether knowledge's constituent beliefs are to be justified by internal criteria of rationality, or made probably true by external relations to their subject matter (*cf.* Grandy [1980]). Kyburg's epistemology recommends internal rationality. But the object of the exercise is external (probable) truth. Since Kyburg's rationality does not deliver that, I see no merit in it.

D. H. MELLOR
University of Cambridge

REFERENCES

GRANDY, RICHARD E. [1980]: 'Ramsey, reliability and knowledge', *Prospects for Pragmatism* (*ed.* D. H. Mellor), pp. 197–210. Cambridge University Press.

MELLOR, D. H. [1980]: 'Consciousness and degrees of belief', *Prospects for Pragmatism* (*ed.* D. H. Mellor), pp. 149–173. Cambridge University Press.

NOZICK, R. [1969]: 'Newcomb's problem and two principles of choice, *Essays in Honor of Carl G. Hempel* (*ed.* N. Rescher), pp. 114–146. Reidel.

MAGEE, BRYAN [1983]: *The Philosophy of Schopenhauer*. Oxford University Press. £17.50. 394 pp.

Schopenhauer's philosophical system is an extension and modification of Kant's transcendental idealism. According to Kant, space, time and causality are subjective ways of ordering phenomena and not objective realities. In his Refutation of Idealism Kant established, against Berkeley, that the subject cannot order its experiences by these subjective forms and categories, or be self-conscious at all, unless the perceived world is empirically real and outside its mind. But hidden behind phenomena are things-in-themselves or noumena which, he says, are the causes of the subject's perceptions but are also necessarily outside human experience. According to Schopenhauer, however, there is only one noumenon, and it is possible to discover on the basis of introspection that this is *Wille*, this being the term he uses for a conjectured impersonal and undifferentiated energy analogous to Spinoza's *conatus*, manifest both in ourselves (in, among other things, acts of will, as the term suggests), and in the physical world. Thus his most important work is entitled *The World as Will and Representation*, or as noumenon and phenomenon.

Kant was clearly concerned with what he regarded as the necessary basis of and limits to perception and knowledge. Magee is a convinced transcendental idealist of the Kantian type who is nevertheless also fascinated by Schopenhauer's colourful central notion of Will. It is after all a propensity for bold speculation which makes German philosophy traditionally both more interesting and more irresponsible than British. German systematic philosophy after Kant, reaching its apogee in Heidegger, is largely built on what Kant called synthetic *a priori* judgements, usually of great generality. These, being neither logical truths nor statements of fact, nor indeed often very precise, invite us to view the world in a certain way for the sake of new insights. According to Schopenhauer the central experience of life was the gaining of new insights, but in his case, as in others, we remain free to reject particular insights as synthetic falsehoods, while no particular rejection requires us to reject the system of which an insight is part if we still find the basic *Weltanschauung* convincing.

Schopenhauer is renowned for his pessimism and misanthropy. According to Magee (p. 14) pessimism is not an essential feature of the system because optimism and pessimism are merely ways of contemplating neutral states of affairs, as illustrated in the following picture: Two men are drinking together. They shoot simultaneous glances at the bottle they are sharing, and one thinks to himself 'Ah good, it's still half full' while the other thinks 'Oh dear, it's half empty'. If this is correct, however, both attitudes are compatible with all states of affairs and a man may say 'I'm an optimist. The bomb will drop tomorrow' without making a joke about its not dropping today. Surely, however, optimism and pessimism are to do with predictions and not states of affairs such as half-empty bottles, an example by which Magee merely distinguishes the gloomy from the sanguine. Schopenhauer's view of life as a wearisome tragedy, while not a consequence of his ontology, in that Will need not have been inherently evil, is still basic to the system to the extent that system-building is persuasive. Man was before birth and will be after death part of the general Will; his sexual activity is a manifestation of the same Will that drives a machine. Above all, man is cruel, although he is capable of compassion on occasion. Compassion is indeed for Schopenhauer our highest moral attribute, and it is in reaction against him that Nietzsche makes it our lowest failing. 'Optimism,' Schopenhauer writes, 'where it is not merely the thoughtless talk of those who harbour nothing but words under their shallow foreheads, seems to me to be not merely an absurd, but also a really *wicked* way of thinking.'

Schopenhauer was a determinist, believing that the mind is no more than 'a lantern illuminating the way of the will'. Magee presents the standard arguments against determinism, but admits that he can make little of Schopenhauer's somewhat obscure views about the freedom of man *qua* noumenal entity which has chosen a certain way of behaviour in the phenomenal world being ultimately restored to him (or it) in death. 'The notion that we have in some intelligible sense chosen to be what we are would meet all the main requirements of Schopenhauer's philosophy,' he writes. 'Unfortunately it cannot be coherently formulated, or so it seems to me' (p. 207). On the contrary, I should say, the notion has been coherently formulated in the existentialist notion of a fundamental project, whereby Sartre claims that Jean Genet decided to be fundamentally a thief. But it cannot be formulated within Schopenhauer's deterministic system of Will. Schopenhauer may also be said to share with existentialists the further view, to which he ought not to be entitled, that man is condemned to freedom *qua* conscious or noumenal being.

Magee writes particularly well about Schopenhauer's aesthetics, the most outstanding feature of which is the doctrine that all arts except music present us with Platonic ideas, which mysteriously mediate between noumenon and phenomena. Music, however, like sexual desire and acts of will in general, allows us to experience noumenal energy direct. This doctrine was later developed independently by Wagner, who doubtless was not enough of a philosopher to appreciate the spuriousness of the claim that music, in being non-representational in the sense that some painting is abstract, is therefore not constructed out of phenomenal sounds, and hence is noumenal in origin. According to Schopenhauer, however, music raises the stage curtain to reveal what is being enacted behind the Veil of Unknowing, even if the only player is undifferentiated energy and really rather insipid. As Magee points out, he does not explain why music delights rather than nauseates, for what we surely should experience in direct contact with the noumenon is disgust or boredom. On the other hand Schopenhauer does associate volition with disgust, but not sex. (Kant's discussion of this topic, according to him, was 'very superficial and without specific knowledge', since had had not met the noumenon in that particular avatar.)

Schopenhauer's admiration for ascetics, a class to which he did not himself belong, is reflected in his view of poetic tragedy as carrying a cathartic power which should seize an audience at the moment when the hero or heroine realises the vanity of human effort. In *Hamlet*, Magee tells us (pp. 179–80), he identified the moment with the words 'There's a special providence in the fall of a sparrow— if it be now, 'tis not to come; if it be to come, it will not be now; if it be not now, yet it will come: the readiness is all', a passage also reminiscent of Heidegger's notion of 'waiting on Being' but the absolute antithesis of the noble commitment to life-enhancing action expressed at the high points of *Faust* by Schopenhauer's early idol, Goethe. This rather facile view of a blue-print to which *all* great tragedy must conform is an example of the systematiser's need to force diverse phenomena under one description and thus devalue art or nature as it really and diversely is, a process perhaps accentuated in the more idiotic straight-jacketing of structuralists.

The use of the notion of Platonic ideas in Schopenhauer's aesthetics and in his system in general is deplored by Magee, who prescribes 'a careful shave with Occam's razor' (p. 259). His other main points against the system in a chapter headed 'Some criticisms and problems', apart from the one about music, concern determinism and Schopenhauer's acceptance of the

Kantian assumption that Newton's laws were immutable and men necessarily so constituted as to perceive a world to which they immutably apply. This latter point, however, surely requires only minor modifications to Kant's system and hardly counts against Schopenhauer at all. What appears to be lacking in Magee's book is an exposition or criticism of the idea of system-building in general, an activity which was clearly easier when men believed that philosophy as an independent discipline, as opposed to philosophy of (say) science or education, was a valid pursuit. Surely only a lunatic would elaborate a new philosophical system using the method of *a priori* synthetic construction in the late twentieth century. To Schopenhauer, however, such an activity seemed the highest intellectual exercise. It appears to me that the attempt to address to intelligent men moral speculation about the place of man in the universal scheme of things is still in order, but that the voices listened to with most respect can no longer be those of philosophers, since philosophy as a distinct discipline is only sustained for sociological reasons. There may even be some reason to continue identifying possible noumena, for the urge to do so is within most thinking men, even if only for Freudian reasons. A modern candidate might be Sheldrake's much ridiculed hypothesis of morphic resonance as the vehicle of so-called formative causation, which is more closely comparable to Will than the physical concept of energy which Magee somewhat otiosely compliments Schopenhauer on foreshadowing. Energy after all manifests itself all the time in gross, measurable ways, which is not a noumenal thing to do, whereas Sheldrake's causation, like Jung's synchronicity principle, is a mysterious formative power hidden behind and shaping observables.

Magee devotes the second half of his book to a series of short chapters about Schopenhauer's influence on other philosophers and writers. His appeal to the creative imagination must usually have been based on a limited knowledge of his system, but among novelists Thomas Mann had a clear knowledge of it. In *Buddenbrooks* Thomas Buddenbrook reads an important chapter of *The World as Will and Representation* at one go before proceeding to compose his will. Mann's choice of this chapter well illustrates the lure of that luxuriant pessimism which Magee takes to be peripheral to the system. It is entitled 'Of death and its relationship to the indestructability of our essential being' and begins thus: 'Death is the real inspiring genius of philosophy. . . . Indeed it would be difficult to philosophize were it not for death. So it will be quite in order for a special consideration of it to find its place here at the head of the last, most serious and most important of our books. . . .' (The whole work is divided into four books.) The mystery of death dissolves thus: 'In appearance, which is mediated by its forms Time and Space, it seems to us . . . that the human individual dies, but the human race remains and lives for ever. But in the essential being of things . . . the whole distinction between the individual and the race disappears.' Intimations of this universality are, I think, aroused in us by music (the language of the Will) heard in certain moods, the effect of merely good music being to console the individual for some psychic defeat and seduce him away from suicidal thoughts, while the greatest music persuades us that the life of the race continues untroubled by individual worries. This phenomenological report, however, is given by a slightly more optimistic witness than Schopenhauer.

The chapter ends in a rhapsodic passage which, not surprisingly, contains several *non sequiturs*: '... from the metaphysical point of view the statements "I die, but the world goes on" and "The world dies, but I go on" are fundamentally one and the same.... Death is the great opportunity to be no longer I. . . . During life the will of man is unfree; his behaviour proceeds with inevitability upon the basis of his unalterable character along the chain of motives.... If man lived for ever, then thanks to the unalterability of his character he would also behave in the same way for ever. So (*demnach*) he must cease to be what he is, in order to be able to emerge from the kernel of his being as something new and other. Thus (*daher*) death dissolves those bonds; the will becomes free again: for (*denn*) freedom lies in the Esse, not the Operari.' We see here that Schopenhauer's 'most important' insights are presented in a way that, *pace* Magee, makes little attempt at argument and relies on intoxication by language.

Magee rightly treats the opinions of great informal thinkers such as Wagner and Mann with profound respect. He is impressed by their belief in a noumenal (as opposed to numinous) reality. It may indeed be that such a belief is essential to creative art for some people, especially Germans. Certainly post-Kantian systematic speculation depends on the hypothesised noumenon. What distinguishes this irresponsible use of a theoretical entity from normal practice, however, is that it is without any observable effects and all on its own is supposed to explain the whole of reality. No consequences of any kind follow about what may actually be expected to happen in our experience. If not Will but (say) Love were postulated as the noumenon it would make no difference at all. Schopenhauer in fact commits precisely the abuse of metaphysics against which Kant had protested by applying the categories and forms through which the Kantian subject constructs the objective world from phenomena to the noumenal doings of Will.

Magee persuades me that I should prefer a noumenal world to exist but not that there is any logical need for it. Crude empiricism's rejection does not lead straight to transcendental idealism. He correctly claims that transcendental idealism anticipated the anti-empiricist discoveries of Piaget and Chomsky, but he ignores or merely dismisses as common sense prejudice one of the most basic genetically programmed unlearned schemata of all. For the belief that the perceived world is adequate to experience, the failure to take the noumenal seriously, seems to be at the heart of experience, as indeed does a disbelief in Platonic ideas. To systematic philosophy the noumenal world is a necessity, since without it its practitioners would have no subject matter but would rather allow us to experience the world as it lies before us, delightful and diverse, but complex as a field of action only in so far as it contains physical difficulties and the necessity to interact with other minds. The belief in this world seems to me to invite the same characterisation as that which Freud gave to romantic love— the psychosis of the sane. In this sense Schopenhauer and his fellow systematizers were necessarily not of sound mind, since if they had not been insane they would have had no subject matter. Magee gives a cogent quasi-psychoanalytic account of Schopenhauer's early life. I hypothesise additionally that the socially incompetent and presumably schizoid Kant and the misanthropic, surely manic-depressive Schopenhauer repressed the problem of other minds and projected their puzzlement about the motives of their fellow men on to the external world. The doctrine of the noumenon, on this theory, is a reflection of the unfathomability of other people. (The compiler of Magee's index, being strictly sane, advises us to seek 'idealism, transcendental' under 'transcendental idealism' but, by a happy accident, omits any entry under the latter.)

As a critical exposition Magee's book does not perhaps displace Patrick Gardiner's patient study. It offers, however, a wider consideration of Schopenhauer's abiding influence in diverse fields and an enthusiastic identification with the transcendental idealist tradition. The overwhelming impression it leaves is as much of the greatness of Kant as of the imaginative power of Schopenhauer. The book is elegantly written, as befits a study of a noted German stylist. British critics, like Magee, tend to see lucidity as essential to good literary style. But Schopenhauer is by no means always as clear as Magee suggests, and there are points at which readers of *The World as Will and Representation* will have difficulty interpreting the text. In presenting his philosophy so concisely Magee does Schopenhauer a greater service than he himself allows.

GORDON BLAIR

RADNITZKY, GERARD, and ANDERSSON, GUNNAR (*eds.*) [1979]: *The Structure and Development of Science*. Boston Studies in the Philosophy of Science, vol. LIX. Reidel. Pp. ix+282.

The editors tell us: 'This book evolved from a workshop on topical problems in philosophy and history of science . . .' (p. ix). The results of this workshop, held in 1975, have appeared in two volumes. The first, by the same editors, is *Progress and Rationality in Science* (volume 58 of the Boston Studies in Philosophy of Science), and the second is the book under review. The first book displays its origin in a workshop. It begins with a 'position paper' by members of the Philosophy Department of the London School of

Economics and is followed, in considerable part, by a critical response to the 'position paper' which in turn is followed by replies and rejoinders. The second book has no such useful structure or unifying themes and belies its workshop origin; instead it is the all too common collection of disparate papers.

The nearest approach to the workshop style of the first volume is, first, a pair of papers by Stegmüller and Kockelmans in which the latter discusses (*per accidens?*) some views of the former and, second, a dialogue constructed by Feyerabend. Stegmüller's contribution, which, we are told, had already appeared in *Theory and Decision*, vol. 9 (1978), is a restatement of the "structuralist" (or non-statement) interpretation of theories and its application to Kuhn's idea of a paradigm and Lakatos' idea of a scientific research programme. Those who have followed Stegmüller's work will find nothing new here; however the paper is a useful succinct summary of his position. Joseph Kockelmans in his 'Reflections on Lakatos' Methodology of Scientific Research Programmes' accepts the new logical analysis of science provided by Stegmüller's approach but claims that there are still some features of the scientific enterprise for which that approach can not account. We are then treated to a sketch of Bachelard's ideal of 'recurrent history' and to the claim, with little supporting argument, that there is a *telos* of the scientific enterprise that only a phenomenological analysis can capture. What the '*telos* of science' amounts to is so unclear that it is hard to determine whether Kockelmans can embrace both the Stegmüller logical analysis of theories and his own phenomenological analysis of that *telos*. For Stegmüller rejects explicitly any teleological approach when he speaks of the 'teleological myth of progress' (p. 171) and of the alleged linear development of science as 'a vestige of teleological phantasy' (p. 175).

Feyerabend's 'Dialogue on Method' is the longest piece in the book—70 pages; being a dialogue perhaps one could charitably view it as a one man 'workshop in progress'. Two characters, prosaically named 'A' and 'B' (though it is clear that A is a Popperian critical rationalist who is the fall-guy for B, Feyerabend himself), slug it out over a number of issues that Feyerabend has discussed elsewhere in his recent books, articles and replies to reviewers. We cover, amongst other things, such topics as: a defence of astrology; the competence of scientists; the value of modern western medicine; modern theories of education and learning; the connections between theatre, poetry and science; problems of observation. Only a part of the dialogue is directly devoted to a criticism of aspects of Popper's theory of scientific method. Though Feyerabend has had a lot to say elsewhere on these topics that deserves close attention, the present dialogue is not the best starting point for an examination of his views. On page 85, concerning A's objections to B's views on the comparative worth of research in cancer and research in alternative forms of medicine, B says: 'First study, then speak. I do not see why I should constantly have to refute opinions whose sole support is the conceit of ignorance.' Good advice! So let us apply it to B's defence of the view that in the story of the Rabbi Akiba, who was said to have ascended through the spheres of the heavens in order to finally face God, there is still a viable sense in which we can speak of spheres and that the story is not wholly unbelievable. B proffers for us the current view that there is a Roche boundary (the distance to which one planet can approach another without either being torn apart) in support of the claim that the Roche boundary 'might well represent the spheres in our stories' (p. 104). In the dialogue A is left perplexed—as well he might be! Such an argument (presented on pp. 102–5) is truly the 'conceit of ignorance' on the part of B. With no adequate response from A the reader's sympathy for him is readily aroused since he has become the sorry figure of a fall-guy for the Feyerabendian B; for as soon as A as much as lifts himself to his knees from the prone position he has occupied with respect to B throughout most of the dialogue, the victorious B pushes A prostrate again at his feet. The supine presentation of the critical rationalist's position and the lack of genuine dialectical engagement between the two protagonists in such encounters as the one just mentioned is a central weakness of the dialogue.

The remaining papers show no sign of having been forged in the heat of a workshop. One of these has also been published previously—this is Max Jammer's 'A Consideration of the Philosophical Implications of the New Physics'. Jammer wishes to deny the somewhat bland doctrine of the philosophical neutrality of science, *viz.*, 'philosophy and science are completely separate and independent intellectual disciplines' (p. 41). The implications, he claims, go both ways; philosophical doctrines permeate the new physics while changing theories in physics have forced changes in the philosophical presuppositions of older physical theory with the consequence that there are new problems with which philosophers must come to terms. The interest in the paper is not so much in its philosophical argument but in the sketches of historical cases which conflict with the neutrality doctrine.

Rescher contributes an essay in erotetics. Will science one day answer all questions or will it produce fresh questions in the course of answering the old? Rescher opts for the latter. But that there is no termination to questions is compatible, he argues, with the claim that there may be questions for which no scientific answer is possible, *i.e.*, there are genuine insolubilia and therefore limitations, in one sense, to our knowledge. Concerning the granddaddy question of them all, 'Why is there anything at all?' Rescher canvasses some possible kinds of answer but breaks no new ground concerning this perennial question. Peter Hodgson, an Oxford physicist and the only practising scientist to make a contribution to the book, has a paper 'Presuppositions and Limits of Science' which by title, and in part by content, indicates that it deals with themes similar to those of Rescher's paper; however neither author dialectically engages the other which is surprising considering the workshop context in which the papers were first presented. Hodgson's paper also ranges over a wide number of other topics such as the definition of science, the social and intellectual conditions that underlie science, regularities in science, scientific techniques and the selection of data for scientific treatment. Unfortunately each topic gets at best superficial treatment.

Lakatos' methodology of scientific research programmes concerns the appraisal of a linear sequence of theories of a given hard core and positive heuristic. Heelan, in his 'The Lattice of Growth in Knowledge', advocates replacing the linear sequence of theories by a lattice of quite different elements, a context of inquiry and a language. Once this is done it is unclear how much of Lakatos' original idea of a scientific research programme survives. In addition, Heelan marries his account to some conservative principles of Heisenberg that Lakatos would surely reject, *viz.*, that there be no proliferation of competing theories and that existing theories be conserved as far as possible (p. 208). Heelan raises the issue of how one might provide a non-reductionist account of the manner in which a number of programmes may be superseded by one progressive programme but the details of this are left to work in progress.

Radnitzky, in a 43-page paper, advises us to give up the search for a justification for holding a theory and to settle for the more realistic rational preference of one theory over another. But what is to count as rationally preferable? A number of Popperian ideas are canvassed (including greater empirical content, degree of corroboration and verisimilitude) as well as Lakatos' theory of method. Nothing new emerges from this and we are left with the conclusion that there may be a number of Popperian and Lakatosian grounds on which we can prefer one theory to another, to which Radnitzky adds a consideration of the degree of importance of the problems that the theory solves or tries to solve. Colin Howson is interested in the problem of how Lakatos' methodology of scientific research programmes can be extended to non-empirical disciplines such as mathemtics. A stumbling block seems to be the notion of empirical content so central to Lakatos' methodology. How can mathematical theories have an analogue to the notion of empirical content for scientific theories? Such an analogue is not to be found in Platonism in mathematics in which there is a realm of independent entities which we intuit in some extra-sensory way. Following some remarks of Gödel, Howson suggests that theorems demonstrated without a set of foundational axioms can constitute the empirical analogue for any new set of foundational axioms which may yield simpler proofs and be abundant in their consequences. Finally Howson suggests that Lakatos' criteria of appraisal, *i.e.*, those of progressing, stagnating and degenerating programmes, can be fruitfully applied to research in mathematics yielding distinctive mathematical research programmes.

One of the editors, Andersson, provides an introduction in which he summarises, at greater length than is possible in a review of this sort, the contents of each paper and somewhat vainly tries to link then together under the book's broad title. Of the two volumes to emerge from the workshop this is the less successful.

ROBERT NOLA
University of Auckland

ROSEVEARE, N. T. [1982]: *Mercury's Perihelion from Le Verrier to Einstein.* Oxford University Press. Hardback £20. Pp. 208.

In September 1859 Le Verrier, director of the Paris Observatory, announced "a grave difficulty, worthy of the attention of astronomers". The orbit of the planet Mercury is not quite a closed curve; its perihelion is known to advance because of the perturbing effect of the other planets. But, according to Le Verrier, the observed effect was an advance of 566 seconds of arc per century, while calculations of the influence of the known masses in the Solar System gave 527 seconds. Le Verrier had noted a similar problem in his provisional theory published in 1843, but at that time he could not rule out an inadequacy in the theoretical analysis. In the intervening years, Le Verrier had successfully predicted the existence of the planet Neptune from anomalies in the motion of Uranus. So by 1859 blame could not be laid on the methods of calculation; there was an anomaly of about 39 seconds per century. Modern analysis gives about 43 seconds. Anomalies exist also for the other planets, but are much smaller and consequently more uncertain.

The explanation of the anomaly was found by Einstein in November 1915. On the basis of what was not quite the completed general theory of relativity he obtained an additional perihelion advance of just 43 seconds, almost exactly equal to the value of 45 seconds for the anomalous advance that Freundlich had given him. This calculation was unaffected by subsequent modifications to the general theory. "For a few days, I was beside myself with joyous excitement", Einstein wrote to Ehrenfest later.

Such is the text book history of the perihelion anomaly. But, as Roseveare shows, it obscures a long history of confusion, speculation and misplaced industry, as well as some mis-observation and professional rivalry, a history that is not without relevance for the philosophy of science. Celestial mechanics is particularly interesting because, contrary to the conception even of some physicists, only the theory itself tells the observer what he sees—apart, of course, at the trivial level of points on the celestial sphere. Even contemporary radar determinations of planetary positions can be interpreted at a fundamental level (*i.e.* to better than the Newtonian approximation) only in terms of parameter fitting—including the planetary positions—to a general relativistic theory (which contains a description of the behaviour of electromagnetic radiation in gravitational fields). The precession of Mercury's perihelion is important since, in contrast to some other problems in celestial mechanics, it is sensitive directly to departures from an exact inverse square law of gravitational force—as indeed Newton himself was aware. In addition, of course, the precession depends on the distribution of planetary, or other, masses. Broadly speaking, therefore, the anomaly elicited two types of response. The 'conservatives' saw it as an indication of further matter in the Solar System: an inner planet, asteroid rings, the matter of the zodiacal light. The 'radicals' saw it as a call for a new theory of gravity to replace Newton's inverse square law.

Roseveare documents in detail the theories and the objections to them. Of the matter theories, an unseen inner planet of sufficient mass was observationally absurd, and matter rings with the necessary large inclination met theoretical objections. On the other hand Seeliger's zodiacal light hypothesis received widespread acceptance, not all of which was withdrawn after 1915 when Freundlich showed it to be quantitatively untenable—a result that Einstein referred to as "forcing an open door". Force laws depending on distance only to other than the inverse square power seem to be fairly easily refuted, but velocity dependent force laws are of rather more interest. Of course, there is a clear possibility of analogy here with electrodynamics one of the fascinating aspects of which Roseveare does not consider. It is in fact possible to extend the analogy between the inverse square laws of gravity and electromagnetism by introducing a gravitational analogue of the magnetic force, and to arrive thereby at a derivation of the relativistic perihelion precession. Indeed, this is just a way of describing the weak field approximation to general relativity, and pursued to its conclusion the argument can be made to yield Einstein's field equations. That in the plethora of modified force laws none came near such a development seems to reflect the unlikely nature of Einstein's equations in the absence of the underlying philosophy.

The transition to general relativity was not therefore the sudden appearance of a challenge to Newtonian universal gravity. There were, literally, hundreds of non-Newtonian gravitation theories (just as there are now non-Einsteinian proposals). On the other hand, Dirac has related how general relativity came as a sudden shock to much of the physics community because gravitation theory was not a central issue in physics at the beginning of this century. During the difficult groping towards a satisfactory theory in the period up to 1915 Einstein expressed the feeling that knowledge of gravity was no more advanced than electromagnetism had been when it consisted only of the Cavendish inverse square law, but it was "remarkable how little my colleagues are susceptible to the inner need for a *real* relativity theory".

In those years there was only one rival theory that was not in conflict with relativity nor contained arbitrary parameters. This is Nordström's ('second') scalar theory. As Roseveare points out, Einstein's arguments against it are all invalid. The theory is peculiar, since electromagnetic energy does not have a gravitational effect, but not illogical. It does, as Einstein repeatedly stressed, predict zero bending of starlight by the Sun, in contrast to the various preliminary versions of general relativity. Nevertheless, it was fortunate for the drama of relativity that early attempts (in 1912 and 1914) to measure the light bending were unsuccessful. For until the final theory of 1915 Einstein's prediction for the bending was a factor two too small. On the other hand Einstein was, in public, completely silent on the perihelion advance predicted by the preliminary theories, since these did not agree with observation. Nevertheless, according to Pais in his recent biography, Einstein considered the incorrect precession according to the Einstein–Grossmann theory (18 seconds per century) as an argument against it. He did not, in public at least, employ the incorrect precession (-7 seconds) against the Nordström theory, although this would have been the only available valid argument against it. Roseveare disavows speculation on the reason. To be fair to Einstein, I do not think the reason is entirely transparent. Einstein was concerned initially not primarily with the particular anomaly in the orbit of Mercury, but with the anomalous position of gravity and non-uniform motion in relativity theory. So long as he was unsure whether he had solved this latter problem a calculation of the perihelion precession would leave him little wiser. On the other hand it followed directly from Einstein's version of the equivalence principle (a rather stronger version than that in Nordström's theory) that light must be bent by gravity, and therefore any theory that might be at least on the right track must, in his view, predict some bending.

In fact, Einstein does not appear to have calculated the perihelion precession on the basis of a 'naive' application of the equivalence principle. Roseveare spends some effort to present an explicit calculation which shows that this gives only two-thirds of that from the full theory, although the method here is really no different from the standard theory. In fact the whole rather extensive discussion of misguided attempts to derive a metric for the Solar System on the basis of the equivalence principle could be dispensed with rather simply. For it is quite impossible to pass from the equivalence principle, which is the statement that the laws of non-gravitational physics are the same in all freely falling frames of reference, to a metric, which provides the information as to which frames are in free fall. Einstein did not use such 'plausibility arguments' because they are fallacious. Of course, it is just this problem that occupied Einstein in the period 1907–15 and which is resolved by the field equations.

Roseveare's book is concerned only rather marginally with more recent developments, but I think it is worthwhile finally to put the perihelion precession of Mercury briefly in its contemporary context. It is now understood that a theory of gravity which satisfies the above (strong) form of the equivalence principle, and in which the laws of physics in free fall are locally consistent with special relativity, must be a 'metric theory'—in essence, must be based on a curved space-time (pseudo-Riemannian) geometry. The Solar System metric for any such theory can then be described by a number of theory-dependent parameters. All recent tests have tended to embrace the general relativistic values of these parameters, with increasing precision, and to rule out rival theories. Nor do we have to rely on the Solar System. The binary pulsar consists of a pulsar, which provides an accurate clock, in an eccentric orbit about another star. At 4.2 degrees per year the predicted periastron advance is in this case no small effect. Observations of this system too, which represent a more severe test of the theory than does Mercury, are in agreement with general relativity.

DEREK RAINE
University of Leicester

WEBSTER, CHARLES [1983]: *From Paracelsus to Newton: magic and the making of modern science*. Cambridge University Press. £12.50. 107 pp. ISBN 0 521 24919 8.

Charles Webster's Eddington lectures given at Cambridge in 1980 are published here with the addition of ample references and some fascinating illustrations. This brief and expensive book contains a masterful survey of the fundamental problem in the historical interpretation of the Scientific Revolution. Since Webster considers the often sterile secondary literature on this topic, and himself offers an interpretation which shifts the ground of the debate, the book is also a masterpiece of clarity and compression. Three areas of natural knowledge are considered: prophecy (which here is taken to include both apocalyptic hermeneutics and astrological cometography), together with spiritual and demonic magic. The geographical scope is also rather limited, since in each case a path is made out which moves from Reformation Germany to Restoration England, with occasional glances at France and Italy. The impact of this thrice-told tale is considerable: Webster argues for a profound continuity between the work of Paracelsus as presented in a remarkable series of ephemeral or substantial texts in the sixteenth century, and, on the other hand, the classical texts of the Scientific Revolution, in which those of Newton and Boyle take pride of place. Secondly, he claims that the 'scientists were dragged along with the tide' in changing attitudes to the occult, and thus that historians cannot appeal, as many have done, to changes in natural knowledge when seeking to explain this primary transformation in the status of magic.

These arguments involve a bold strategy, in which it is necessary to reassess the place of Paracelsus, a process which Webster has also touched on in his response to Debus's *Chemical Philosophy* and in a recent paper on the Paracelsian theory of mining and of mountain spirits. Thus Webster gives a substantial and sympathetic account of the prophetic texts of Paracelsus, virtually the only ones published in his lifetime, and then shows how these texts could come to exercise an important influence on subsequent astronomical and interpretative developments. Furthermore, Webster must show the disparate character of the mechanical philosophy, unpacking that term to display the interest of so-called mechanists in the realm of spirit, and showing how natural philosophy in the later seventeenth century by no means limited itself to pure matter in motion. In these attempts he recalls his previous claims that 'the mechanical philosophers' of the generation of Henry Power and others 'had no firmer foundation for their explanatory devices than their Peripatetic critics', and so calls into question some recent comments, such as those of Cantor and Hodge, that men like Boyle saw themselves 'separated by vast metaphysical gulfs from any Platonists or equally any Paracelsans'. In fact, it is necessary to break with much of recent scholarship, both that of Yates, for whom the occult philosophy influenced later science in its rarified Hermetic form, and that of Keith Thomas, for whom the 'triumph of the mechanical philosophy' was the 'essence' of the scientific revolution, and who notes the 'end of the animistic conception universe' while labelling Newton's alchemy as 'cranky'. Most interestingly, Webster's book prompts a question in historiography: why has it been so important for historians of science to pose this problem in terms of some definitive and progressive break with all previous thought styles, and why has it proved impossible to locate that break in an unambiguous fashion?

The answer implied here is a radical one. Historians have viewed the task of the history of science as the catalogue of progress, and have then identified such progress with the work of a set of privileged individuals. Terms such as 'occultism' or 'mechanism' are then defined in a manner which dovorces them from specific contexts of use, but links them with those heroes who can be assigned to specific schools: thus Paracelsus becomes occultism incarnate, while Boyle stands for the mechanical philosophy in action. It is then easy to show that these taxonomies break down. Boyle's encouragement of the collection of witch testimony, his 'apology for astrology' sent to Hartlib and published in the *General History of Air*, or his use of terms such as 'supramechanical' in the posthumous part of the *Christian Virtuoso*, can all be used to break down the mechanical label. Similarly, a hero of iatromechanism such as John Mayow introduced his *Tractatus quinque* with an explicitly Paracelsian commentary on 'the nitro-aerial Mercury', wrote of the 'sensitive soul' as a 'divine aura endowed with sense from the first creation and co-extensive with the whole world', yet condemned the concept of a lucid soul as 'vain and fanatical' both in anatomy and in religion. The literature in which apparently occultist authors are revealed as contributors to progress is also considerable and equally easy to exploit to this purpose. But this campaign has missed an important point: the issue cannot concentrate on the evaluation of rival claimants to progress, but must instead rely on the detailed assessment of usage in specific contests in natural philosophy.

This argument is exemplified best in the cases Webster considers here. Recent work by R. Mandrou and Erik Midelfort on witchcraft in France and Germany has claimed that decline in witch belief depended on changes in legal and political power, and specifically on a critique of the capacity of courts in such cases. Here 'occultism' becomes a category in the possible scope of certain forms of authority and power. For Boyle, 'the invisible marks of witches' would 'help to enlarge the somewhat too narrow conceptions men are wont to have of the amplitude of the works of God': apparently naturalistic accounts of possession or of cure, such as those offered by Henry More or his allies, by no means excluded appeals to 'occult' spirits, nor, alternatively, did attacks on witch beliefs, such as that by John Wagstaffe referred to by Charles Webster, exclude spirit belief. Wagstaffe wrote that while he could 'firmly believe that there are many thousands of spirits made of an incorporeal matter too fine to be perceived by the senses of man', nevertheless he denied that spirits could 'play mad pranks amongst us' merely 'upon the account of a contract made with man or woman'. Finally, for the arch-mechanist Thomas Hobbes, witch prosecutions were highly desirable, just 'for the false belief they have that they can do such mischief, joined with their purpose to do it if they can'. It would be impossible to divide these views on the lines of occultists and mechanists: as Hutchison argued recently, it was their preparedness to admit previously 'occult' qualities which marked the work of seventeenth-century natural philosophers. Here the metaphysics of witchcraft must be disaggregated, and the interests at work in the deployment of abusive terms must be historically located.

Finally, therefore, Webster's book is welcome for its provision of evidence for the construction of new terms of debate in sciences. Here is the background to the detailed revisions to Newton's General Scholium and to his responses to attacks from Leibniz over the character of occult qualities and the revival of 'great scholastic monstrosities'. Webster shows how such terms came to be valuable in such contests: he displays in some detail those aspects of the work of experimental natural philosophy which could betray their true genealogy. In so doing, he seems to have followed Eddington's own intention to formulate a description of science to which scientists 'stand committed by their practice', rather than by their apologetics.

SIMON SCHAFFER
Imperial College, London

194 *Brit. J. Phil. Sci.* 35 (1984), 194–199 *Printed in Great Britain*

Recent Publications

ACHINSTEIN, PETER (*ed.*) [1983]: *The Concept of Evidence.* Oxford University Press. Paperback £5.50. 0-19-875062-5. Pp. 182.

ARIS, RUTHERFORD, DAVIS, H. TED and STUEWER, ROGER H. (*eds.*) [1983]: *Springs of Creativity. Essays on Founders of Modern Science.* University of Minnesota Press. Hardback $32.50. 0-8166-1087-8. Pp. 342.

ARMSTRONG, D. M. [1983]: *What is a Law of Nature?* Cambridge University Press. Hardback £17.50. $29.95. 0-521-25343-8. Pp. 180.

ASQUITH, P. D. and NICKLES, T. (*eds.*) [1983]: *PSA 1982 Vol. Two: Symposia. Proceedings of the 1982 Biennial Meeting of the Philosophy of Science Association.* Philosophy of Science Association, Michigan. Hardback $25.00. 0-917586-19-0. Pp. xxiv + 730.

ATTFIELD, ROBIN [1983]: *Ethics of Environmental Concern.* Basil Blackwell. Hardback £15.50. 0-631-13137-X. Pp. 220.

AYER, A. J. [1984]: *Philosophy in the Twentieth Century.* Counterpoint, Unwin Paperbacks. Paperback £3.95. 0-04-100044-7. Pp. viii + 280.

BENACERRAF, PAUL and PUTNAM, HILARY (*eds.*) [1984]: *Philosophy of mathematics. Selected readings. Second edition.* Cambridge University Press. Hardback £30.00. $39.50. 0-521-22796-8. Paperback £10.95. $18.95. 0-521-29648-X. Pp. viii + 600.

BERNARD, THOMAS J. [1983]: *The Consensus-Conflict Debate: Form and Content in Social Theories.* Columbia University Press. Hardback $36.50. 0-231-05670-2. Paperback $18.50. Pp. 229.

BERNSTEIN, RICHARD J. [1984]: *Beyond Objectivism and Relativism: Science, Hermeneutics, and Praxis.* Basil Blackwell. Hardback £19.50. 0-631-13429-8. Paperback £7.50. 0-631-13437-9. Pp. xix + 284.

BICKHARD, MARK H. and RICHIE, MICHAEL D. [1983]: *On the Nature of Representation: A Case Study of James Gibson's Theory of Perception.* Praeger, N.Y. Hardback. $23.95. 0-03-069526-0. Pp. xiii + 107.

BLACKBURN, SIMON [1984]: *Spreading the Word: Groundings in the Philosophy of Language.* Clarendon Press, Oxford. Hardback £16.00. 0-19-824650-1. Paperback £6.95. 0-19-824651-X. Pp. xi + 368.

BLACKMORE, JOHN and SEXL, ROMAN (*eds.*) [1983]: *Ludwig Boltzmann Gesamtausgabe: Ausgewählte Abhandlungen der internationalen Tagung Wien 1981.* Friedr. Vieweg & Sohn. Hardback DM 72.00. 3-528-08557-6. Pp. 368.

BLOOR, DAVID [1984]: *Wittgenstein. A Social Theory of Knowledge.* Macmillan Press. Hardback £15.00. 0-333-30016-5. Paperback £5.95. 0-333-30017-3. Pp. 213.

BREGER, HERBERT [1982]: *Die Natur als arbeitende Maschine.* Campus Forschung, Frankfurt. Paperback DM 46.00. 3-593-33082-2. Pp. 252.

BROOKS, JOHN LANGDON [1984]: *Just Before the Origin: Alfred Russel Wallace's Theory of Evolution.* Columbia University Press. Hardback $39.00. 0-231-05676-1. Pp. 284.

BUNGE, MARIO [1983]: *Treatise on Basic Philosophy Volume 5. Exploring the World.* D. Reidel. Hardback Dfl. 160.00. $64.00. 90-277-1511-4. Paperback Dfl. 85.00. $34.00. 90-277-1523-8. Pp. 404.

BUNGE, MARIO [1983]: *Treatise on Basic Philosophy Volume 6. Understanding the World.* D. Reidel. Hardback Dfl. 125.00. $50.00. 90-277-1634-X. Paperback Dfl. 70.00. $28.00. 90-277-1635-8. Pp. 296.

CALDER, NIGEL [1984]: *Timescale: An Atlas of the Fourth Dimension.* Chatto & Windus, The Hogarth Press. Hardback £12.95. 0-7011-3925-0. Pp. 288.

CAWS, PETER [1984]: *Sartre (series: The Arguments of the Philosophers).* Routledge & Kegan Paul. Paperback £6.95. 0-7102-0233-4. Pp. xiii +210.

COSLETT, TESS (*ed.*) [1984]: *Science and Religion in the Nineteenth Century.* Cambridge University Press. Hardback £22.50. $39.50. 0-521-24402-1. Paperback £7.95. $14.95. 0-521-28668-9. Pp. 249.

DALLA CHIARA, MARIA LUISA *et al.* [1983]: *Logic in the 20th Century: A Series of Papers on the Present State and Tendencies of Studies.* Scientia, Milan. Paperback L. 35,000. Pp. 246.

DEUTSCHER, MAX [1983]: *Subjecting and Objecting.* Basil Blackwell. Hardback £17.50. 0-631-13404-2. Pp. 281.

DINER, S. *et al.* (*eds.*) [1984]: *The Wave-Particle Dualism. A Tribute to Louis de Broglie on his 90th Birthday.* D. Reidel. Hardback Dfl. 190.00. $76.00. 90-277-1664-1. Pp. 566.

DURBIN, PAUL T. and RAPP, FRIEDRICH (*eds.*) [1983]: *Philosophy and Technology.* D. Reidel. Hardback Dfl. 135.00. $59.00. 90-277-1576-9. Pp. xiv+343.

DVORAK, JOHANN [1981]: *Edgar Zilsel und die Einheit der Erkenntnis.* Löcker Verlag. 3-85409-019-6. Pp. 172.

EARMAN, JOHN (*ed.*) [1984]: *Testing Scientific Theories. Minnesota Studies in the Philosophy of Science, Volume X.* University of Minnesota Press. Hardback $39.50. 0-8166-1158-0. Paperback $16.95. 0-8166-1159-9. Pp. viii+484.

ELLIOT, ROBERT and GARE, ARRAN [1983]: *Environmental Philosophy. A Collection of Readings.* Open University Educational Enterprises. Paperback £8.95. 0-335-10407-X. Pp. xiv+303.

ERDEI, LÁSZLÓ [1983]: *Der Schluss: Die Dialektik des Schlusses.* Akadémiai Kiadó, Budapest. Hardback Ł 9.90. 963-05-3191-7. Pp. 287.

FARJOUN, EMMANUEL and MACHOVER, MOSHÉ [1983]: *Laws of Chaos. A Probabilistic Approach to Political Economy.* Verso Editions and NLB, London. Hardback £25.00. 0-86091-058-X. Paperback £8.95. 0-86091-768-1. Pp. 264.

FORMENT, EUDALDO [1983]: *Persona Y Modo Substancial.* Promociones

Publicaciones Universitarias, Barcelona. Paperback. 84-86-130-23-9. Pp. 421.

FORMENT, EUDALDO [1983]: *Ser Y Persona*. Edicions de la Universitat de Barcelona. Paperback. 84-7528-104-4. Pp. 523.

FRENCH, PETER A., EUHLING, JR. THEODORE E., WETTSTEIN, HOWARD K. (*eds.*) [1983]: *Midwest Studies in Philosophy Vol. VIII: Contemporary Perspectives on the History of Philosophy*. University of Minnesota Press. Hardback $45.00. 0-8166-1207-2. Paperback $18.95. 0-8166-1212-9. Pp. 539.

FRIEDMAN, MICHAEL [1983]: *Foundations of Space-Time Theories*. Princeton University Press. Hardback £30.40. 0-691-07239-6. Pp. xvi +385.

FRONGIA, GUIDO [1983]: *Wittgenstein Regole e Sistema*. Franco Angeli Editore, Milan. Paperback L. 22,000. Pp. 268.

GARDEN, RACHEL WALLACE [1983]: *Modern Logic & Quantum Mechanics*. Adam Hilger Ltd. Hardback £18.00. 0-85274-457-9. Pp. 176.

GÄRDENFORS, P., HANSSON, B. and SAHLIN, N.-E. (*eds.*) [1983]: *Evidentiary Value: Essays Dedicated to Sören Halldén on his Sixtieth Birthday*. Library of Theoria, Sweden. Paperback. 91-40-04873-X. Pp. 120.

GILBERT, NIGEL G. and MULKAY, MICHAEL [1984]: *Opening Pandora's Box. A sociological analysis of scientists' discourse*. Cambridge University Press. Hardback £22.50. $37.50. 0-521-25418-3. Paperback £7.95. $12.95. 0-521-27430-3. Pp. viii+202.

GOOD, I. J. [1983]: Good Thinking. *The Foundations of Probability and Its Applications*. University of Minnesota Press. Hardback $35.00. 0-8166-1141-6. Paperback $14.95. 0-8166-1142-4. Pp. xviii+332.

GOTTINGER, HANS W. [1983]: *Coping with Complexity. Perspectives for Economics, Management and Social Sciences*. D. Reidel. Hardback Dfl. 100.00. $44.00. 90-277-1510-6. Pp. xv+224.

GRAMMATICAKIS, PANOS [1983]: *Pragmatisme et Science*. Albert Blanchard, Paris. Paperback. 2-85367-035-X. Pp. 91.

GRENE, MARJORIE (*ed.*) [1983]: *Dimensions of Darwinism: Themes and Counterthemes in Twentieth-Century Evolutionary Theory*. Cambridge University Press. Hardback £15.00. $29.95. 0-521-25408-6. Pp. 336.

GROSSMAN, REINHARDT [1983]: *The Categorial Structure of the World*. Indiana University Press. Hardback $25.00. 0-253-31324-4. Pp. 431.

HABERMAS, JÜRGEN [1983]: *Philosophical-Political Profiles*. Heinemann Educational Books. Hardback £16.50. 0-435-82015-X. Pp. xxv+211.

HACKING, IAN [1983]: *Representing and intervening*. Cambridge University Press. Hardback £20.00. $37.50. 0-521-23829-3. Paperback £5.95. $10.95. 0-521-28246-2. Pp. xv+287.

HARE, F. K. (*ed.*) [1983]: *The Experiment of Life: Science and Religion*. University of Toronto Press. Hardback £21.25. 0-8020-2486-6. Paperback £6.75. 0-8020-6506-6. Pp. xvii+185.

HARRIS, ERROL E. [1983]: *An Interpretation of the Logic of Hegel.* University Press of America. Paperback. 0-8191-3544-5. Pp. xiii + 346.

HEELAN, PATRICK A. [1983]: *Space-Perception and the Philosophy of Science.* University of California Press. Hardback £24.25. $37.00. 520-04611-0. Pp. xiv + 383.

HOME, R. W. (*ed.*) [1983]: *Science Under Scrutiny. The Place of History and Philosophy of Science.* D. Reidel. Hardback Dfl. 90.00. $36.00. 90-277-1602-1. Pp. xviii + 182.

HUDSON, W. D. [1984]: *Modern Moral Philosophy.* Macmillan Press. Paperback £6.95. 0-333-35773-6. Pp. xviii + 433.

HUMPHREY, NICHOLAS [1983]: *Consciousness Regained.* Oxford University Press. Hardback £12.95. 0-19-217732-X. Pp. 222.

JAKI, STANLEY L. [1983]: *Angels, Apes and Men.* Sherwood Sugden & Co. and The Handel Press. Paperback £4.75. 0-89385-017-9. Pp. 124.

JEFFREY, RICHARD C. [1983]: *The Logic of Decision.* The University of Chicago Press. Hardback £16.00. 0-226-39581-2. Pp. 231.

KANT, IMMANUEL [1982]: *Critique of Pure Reason. English Translation. Introduction and Glossary by Wolfgang Schwarz.* Scientia Verlag. Hardback DM 98.00. 3-511-09160-8. Pp. xxxvi + 281.

KIRK, G. S., RAVEN, J. E. and SCHOFIELD, M. [1983]: *The Presocratic Philosophers.* Cambridge University Press. Hardback £30.00. 0-521-25444-2. Paperback £10.95. 0-521-27455-9. Pp. 501.

KITCHER, PHILIP [1983]: *Abusing Science—The Case Against Creationism.* Open University Press. Paperback £6.95. 0-335-10199-2. Pp. 213.

KITCHER, PHILIP [1983]: *The Nature of Mathematical Knowledge.* Oxford University Press. Hardback £15.00. 0-19-503149-0. Pp. 287.

KUTSCHMANN, WERNER [1983]: *Die Newtonsche Kraft: Metamorphose Eines Wissenschaftlichen Begriffs.* Franz Steiner Verlag GMBH. Paperback DM 56.00. 3-515-03727-6. Pp. 177.

LAMBERT, KAREL [1983]: *Meinong and the Principle of Independence.* Cambridge University Press. Hardback £19.50. $34.50. 0-521-25085-4. Paperback £6.50. $11.95. 0-521-27199-1. Pp. xvi + 175.

LEWIS, DAVID [1983]: *Philosophical Papers. Volume I.* Clarendon Press, Oxford. Hardback £20.00. 0-19-503203-9. Paperback £8.95. 0-19-503204-7. Pp. 285.

LUKÁCS, JÓZSEF and TÖKEI, FERENC (*eds.*) [1983]: *Philosophy and Culture: Studies from Hungary.* Akadémiai Kiadó. Hardback £10.50. 963-05-3449-5. Pp. xii + 368.

MAZE, J. R. [1983]: *The Meaning of Behaviour.* George Allen & Unwin. Hardback £15.00. 0-04-150081-4. Pp. 186.

MCGUIRE, J. E. and TAMNY, MARTIN [1983]: *Certain Philosophical Questions: Newton's Trinity Notebook.* Cambridge University Press. Hardback £45.00. 0-521-23164-7. Pp. xii + 519.

MEDAWAR, PETER [1984]: *Pluto's Republic.* Oxford University Press. Paperback £4.95. 0-19-283139-2. Pp. 351.

MERLEAU-PONTY, JACQUES [1983]: *La Science de L'Universe à L'Âge du Positivisme*. J. Vrin, Paris. Paperback 270F. 2-7116-0828-X. Pp. 367.

MILLER, DAVID (*ed.*) [1983]: *A Pocket Popper*. Fontana Paperbacks. £4.95. 0-00-636414-4. Pp. 479.

MONTI, MARIA TERESA [1983]: *Catalogo del Fondo Haller della Biblioteca Nazionale Braidense di Milano. Parte I: Libri—Vol. I: A–F*. Paperback L. 35,000. Pp. 608.

MUKERJI, CHANDRA [1983]: *From Graven Images: Patterns of Modern Materialism*. Columbia University Press. Hardback $39.00. 0-231-05266-2. Paperback $16.00. 0-23105167-0. Pp. 329.

PAZUKHIN, ROSTISLAV [1983]: *Linguistic Necessity and Linguistic Theory*. Uniwersytet Slaski, Katowice. Paperback zl. 95. 83-00-00774-1. Pp. 127.

POLKINGHORNE, DONALD [1983]: *Methodology for the Human Sciences*. State University of New York Press. Hardback £28.50. 0-87395-663-X. Paperback £8.45. 0-87395-664-8. Pp. 349.

RAVETZ, JEROME R. [1983]: *Scientific Knowledge and its Social Problems*. Oxford University Press, N.Y. Paperback $9.95. 0-19-519721-6. Pp. 449.

REHBOCK, PHILIP F. [1983]: *The Philosophical Naturalists. Themes in Early Nineteenth-Century British Biology*. University of Wisconsin Press. Hardback £25.50. 0-299-09430-8. Pp. 281.

RESCHER, NICHOLAS (*ed.*) [1983]: *The Limits of Lawfulness*. University Press of America. Hardback $18.75. 0-8191-3176-8. Paperback $8.25. 0-8191-3177-6. Pp. 111.

RESTIVO, SAL [1983]: *The Social Relations of Physics, Mysticism, and Mathematics*. D. Reidel. Hardback Dfl. 114.00. $49.50. 90-277-1536-X. Pp. 309.

RICHARDS, STEWART [1983]: *Philosophy and Sociology of Science: An Introduction*. Basil Blackwell. Hardback £17.50. 0-631-13411-5. Paperback £5.95. 0-631-13414-X. Pp. viii+210.

ROSSVAER, VIGGO [1983]: *The Laborious Game: A Study of Plato's Parmenides*. Universitetsforlaget, Norway. Paperback. 82-00-06531-6. Pp. 87.

RUSSELL, BERTRAND [1984]: *A History of Western Philosophy*. Counterpoint, Unwin Paperbacks. Paperback £4.95. 0-04-100045-5. Pp. 842.

RUSSELL, BERTRAND (*ed.* by Kenneth Blackwell *et al.*). [1983]: *The Collected Papers of Bertrand Russell: Volume One. Cambridge Essays 1888–99*. George Allen & Unwin. Hardback £48.00. 0-04-920067-4. Pp. xxxiv+554.

SARKAR, HUSAIN [1983]: *A Theory of Method*. University of California Press. Hardback $33.75. 0-520-04730-3. Pp. xvii+229.

SCHÄFER, WOLF (*ed.*) [1983]: *Finalization in Science. The Social

Orientation of Scientific Progress. D. Reidel. Hardback Dfl. 135.00. $59.00. 90-277-1549-1. Pp. xviii+315.

SCHICK, FREDERIC [1984]: *Having Reasons. An Essay on Rationality and Sociality.* Princeton University Press. Hardback £19.50. 0-691-07280-9. Paperback £6.90. 0. 691-02029-9. Pp. 158.

SINGER, PETER [1983]: *The Expanding Circle: Ethics and Sociobiology.* Oxford University Press. Paperback £3.95. 0-19-283038-4. Pp. 190.

SONTAG, SUSAN (*ed.*) [1983]: *Barthes: Selected Writings.* Fontana Paperbacks. Paperback £4.95. 0-00-636645-7. Pp. 495.

SORABJI, RICHARD [1983]: *Time, Creation & The Continuum.* Gerald Duckworth & Co. Hardback £29.50. 0-7156-1693-5. Pp. xviii+473.

STAATS, ARTHUR W. [1983]: *Psychology's Crisis of Disunity.* Praeger, N.Y. Hardback $39.95. 0-03-063289-7. Pp. x+391.

STADLER, FRIEDRICH [1982]: *Vom Positivismus zur "Wissenschaftlichen Weltauffassung".* Löcker Verlag. Paperback. 3-85409-038-2. Pp. 349.

STOCKMAN, NORMAN [1983]: *Antipositivist Theories of the Sciences.* D. Reidel. Hardback Dfl. 115.00. $49.95. 90-277-1567-X. Pp. x+284.

TAGLIACOZZO, GIORGIO (*ed.*) [1983]: *Vico and Marx: Affinities and Contrasts.* Macmillan Press. Hardback £30.00. 0-333-35908-9. Pp. 438.

TRAGESSER, ROBERT S. [1984]: *Husserl and Realism in Logic and Mathematics.* Cambridge University Press. Hardback £17.50. $29.50. 0-521-24297-5. Paperback £5.95. $10.95. 0-521-28587-9. Pp. xvii +126.

TURNER, BRYAN S. [1983]: *Religion and Social Theory: A Materialist Perspective.* Heinemann Educational Books. Hardback £15.00. 0-435-82894-0. Paperback £5.95. 0-435-82895-9. Pp. 264.

VON SCHILCHER, FLORIAN and TENNANT, NEIL [1984]: *Philosophy, Evolution and Human Nature.* Routledge & Kegan Paul. Hardback £15.95. 0-7100-9767-0. Pp. 283.

WHEATON, BRUCE R. [1983]: *The Tiger and the Shark: Empirical Roots of Wave-Particle Dualism.* Cambridge University Press. Hardback £22.50. $39.50. 0-521-25098-6. Pp. 355.

WUKETITS, FRANZ M. (*ed.*) [1983]: *Concepts and Approaches in Evolutionary Epistemology.* D. Reidel. Hardback Dfl. £135.00. $52.00. 90-277-1577-7. Pp. xiii+320.

YATES, FRANCES A. [1983]: *Renaissance and Reform: The Italian Contribution* (*Collected Essays Vol. II*). Routledge & Kegan Paul. Hardback £15.95. 0-7100-9530-9. Pp. xii+273.

YATES, FRANCES A. [1983]: *The Occult Philosophy in the Elizabethan Age.* Ark Paperbacks, Routledge & Kegan Paul. Paperback £2.95. 0-7448-0001-3. Pp. 217.

ANNOUNCEMENT

THE SOCIETY FOR THE HUMANITIES
CORNELL UNIVERSITY

Junior Postdoctoral Fellowships for the 1985–86 academic year. Deadline 1 November 1984. Ph.D. by time of application and one or more years of college teaching. Focal Theme: Non-European Traditions in Western Civilization. Further information available from Jonathan Culler, Director, The Society for the Humanities, 27 East Avenue, Ithaca, New York 14853, U.S.A.

HISTORY OF SCIENCE

is the only review of literature and research in the history of science, medicine and technology in its intellectual and social context.

FOUNDED IN 1962 AS AN ANNUAL VOLUME, *History of Science* discusses problems and sources for future research in the history of science, medicine and technology, and surveys recent publications. It now appears each year in quarterly parts of about 112 pages each.

VOLUME 22 (1984) will include

PAOLO CASINI *Newton: The Classical Scholia*

JAMIE C. KASSLER *Man – A Musical Instrument*

SIMON SCHAFFER *Priestley's Questions: An Historiographic Survey*

MICHAEL FORES *Constructed Science and the Seventeenth-century 'Revolution'*

SYLVANA P. TOMASELLI *The First Person: Descartes, Locke and Mind-Body Dualism*

WILLIAM R. WOODWARD *Committed History of Social Science in the Two Germanies*

JOHN V. PICKSTONE *Ferriar's Fever to Kay's Cholera*

GLYN PARRY *Hermeticism and Protestantism*

ALEX DOLBY *Thermochemistry versus Thermodynamics*

STEVEN SHAPIN and SIMON SCHAFFER *Making Newton*

WILLIAM EAMON *The Advent of Printing and Experimental Science in the Sixteenth Century*

together with other articles, reviews, and notices of books.

THE ANNUAL SUBSCRIPTION is $75 post free in the Americas and Japan, £35 elsewhere ($40 or £18.50 to direct private subscribers). New subscribers may purchase the first twenty-one volumes (bound in hard-cover) for $450/£225 post free, *including* the subscription to Volume 22 (1984). Write to: Science History Publications Ltd, Halfpenny Furze, Mill Lane, Chalfont St Giles, Bucks HP8 4NR, U.K.

Nature and System

Philosophical Studies of Natural and Artificial Systems

Editorial Aim: *Nature and System* seeks to continue the theorizing spirit of traditional philosophy of nature within a contemporary scientific context, as such it encourages the reflective investigation of philosophic foundations and presuppositions as well as the creative synthesis of new ideas and models.

Vol. VI (1984) to include:

Martin V. Curd, *Kuhn, Scientific Revolutions, and the Copernican Revolution;* **Jonathan Jacobs,** *Teleology and Essence: An Account of the Nature of Organisms;* **A. Olding,** *On Theories of Evolution and Their Several Natures;* **V. J. A. Novak** and **P. T. Saunders,** *Some Important Aspects of Evolutionary Biology in Socialist Countries;* **V. J. A. Novak,** *An Outline of the Principle of Sociogenesis;* **Jacques Rozenberg,** *Space and Life;* **Walter Rubenstein,** *A Neuroelectromagnetic Model of Observational Consciousness;* Survey of Scientific Periodicals.

Subscription Rates: Individuals ($11 00); Institutions ($14.50). Single issues available to individuals for $3.00. Foreign postage: Add 25 cents for each issue; $1.00 for annual subscription.

All correspondence, manuscripts, and subscription orders to:

P.O. BOX 3368, TUCSON, AZ. 85722. USA

Editor
J L Esposito
Associate Editors
H. C Byerly
K L. Ketner
Advisory Board
J Agassi
F J. Ayala
R B. Brandom
G. Buchdahl
W. Buckley
M Bunge
A W Burks
R E Butts
R. Cappelletti
K. G Denbigh
C Eisele
M. Grene
J Lachs
P T. Landsberg
E Laszlo
E. McMullin
J Mittelstrass
D Nauta Jr
A. Polikarov
J Ransdell
N. Rescher
S Rosenthal
P T Saunders

RATIO

edited by
MARTIN HOLLIS

Volume XXVI (1984) **June**

GEORGE SCHLESINGER	Possible Worlds and the Mystery of Existence
MARIE McGINN	Kripke on Wittgenstein's Sceptical Problem
N.M.L. NATHAN	Social Choice Theory versus The People
NICHOLAS UNWIN	Substance, Essence and Conceptualism
JONATHAN HARRISON	The Truth of Statements about Knowledge, and the Propriety of making them
MARTIN CURD	Showing and Telling: Can the difference between right and left be explained in words?
J.E. TILES	On a New Methodological Individualism
PHILIP PETTIT	In Defence of 'A New Methodological Individualism': Reply to J.E. Tiles

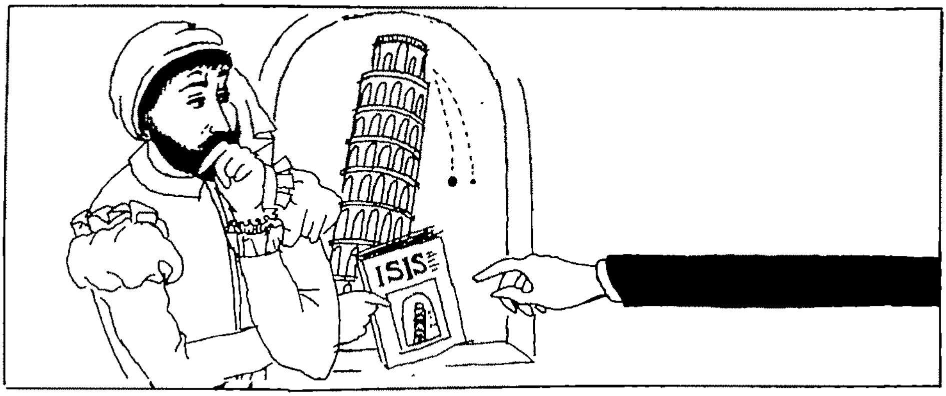
Subscriptions 1984
Institutions £22.95 (UK) £28.20 (overseas) $59.96 (USA) $74.00 (Canada)
Individuals £15.50 (UK) £19 00 (overseas) $42.00 (USA) $52 00 (Canada)

Orders to: Journals Department, Basil Blackwell,
108 Cowley Road, Oxford OX4 1JF, England.

The German edition of *Ratio* is published by FELIX MEINER VERLAG,
Richardstrasse 47, 2000 Hamburg 76, West Germany.

Published by BASIL BLACKWELL, Oxford

ARE YOU STILL READING SOMEONE ELSE'S COPY OF ISIS?



IF SO, now is the time to enter your own subscription. **Isis**, the official journal of the History of Science Society, is the leading journal in the field.

Isis keeps over 3,500 subscribers in nearly fifty countries up to date on all developments in the history of science with articles, critiques, and documents and translations. Along with these, its notes and correspondence and news of the profession provide useful information to professionals, educators, scholars, and graduate students.

Lively essay reviews and over 200 book reviews a year cover every specialty in the history of science, technology, and medicine.

In addition to your four quarterly issues of **Isis** you will also receive:

■ Membership in the History of Science Society.

■ The annual *Critical Bibliography*, listing over 3,000 publications in the history of science, technology and medicine from the preceding year.

■ The triennial *Guide to The History of Science*, containing a directory of members and information on scholarly programs, societies, and over ninety journals in the field. The 1983 edition is now available.

■ The quarterly *Newsletter*, providing current news of the profession, including employment opportunities and approaching meetings.

ISIS

Isis Publication Office, Dept.
University of Pennsylvania
215 South 34th St./D6
Philadelphia, Pa. 19104

YES! Please send me **Isis** for the calendar year(s) 1984 ________ and 1985 ________
$29 for one year ($16 for students). $54 for two years ($29 for students).

☐ Check enclosed ☐ Bill me. (Issues sent on receipt of payment.)

NAME ____________________

ADDRESS ____________________

CITY STATE ZIP PHONE ( )

THEORIA

A Swedish journal of philosophy
Edited by Peter Gärdenfors, Bengt Hansson
and Krister Segerberg

Volume 49, part 1
SPECIAL ISSUE ON EVOLUTIONARY EPISTEMOLOGY

A. Olding Biology and knowledge
Andrew J. Clark Meaning and evolutionary epistemology
Neil Tennant In defence of evolutionary epistemology

THEORIA, which invites contributions in all fields of analytic philosophy, usually publishes three issues a year. Subscriptions and requests for back numbers should be sent to *Editor,* THEORIA, *Filosofiska institutionen, Kungshuset i Lundagård, S-22350 Lund, Sweden.* Subscription for one year is SwCr 60:– or U.S. $ 16.–, for three years SwCr 160:– or U.S. $ 40.–.

AUSTRALASIAN JOURNAL OF PHILOSOPHY

Contents of Vol. 62, No. 1, March 1984

EDITOR: BRIAN ELLIS

ARTICLES

Frederick F. Schmitt	Reliability, Objectivity and The Background of Justification
John Tienson	Hesperus and Phosphorus I
Steven Luper-Foy	The Epistemic Predicament: Knowledge Nozickian Tracking, and Scepticism

SYMPOSIUM

Brian Ellis	Two Theories of Indicative Conditionals
Frank Jackson	Two Theories of Indicative Conditionals: Reply to Brian Ellis

DISCUSSION

Anthony Appiah	Jackson on the Material Conditional

BOOKS RECEIVED, NOTES AND NEWS, INDEX

ANNUAL SUBSCRIPTIONS: Individuals A$16 (US$18 or Stg£9); Institutions A$32, (US$36 or Stg£17). All correspondence should be addressed to the Editor or Business Manager, Department of Philosophy, La Trobe University, Bundoora, Victoria, 3083, Australia.

PUBLISHED QUARTERLY BY
THE AUSTRALASIAN ASSOCIATION OF PHILOSOPHY

EDITORIAL BOARD

R. B. Braithwaite
G. Buchdahl
L. J. Cohen
Sir John Eccles
M. B. Hesse
Sir Harold Jeffreys
C. W. Kilmister
W. C. Kneale
S. Körner
D. J. O'Connor
Sir Karl Popper
H. R. Post
J. W. N. Watkins
G. J. Whitrow

EDITORIAL ADVISORY PANEL

The Editor is helped by members of the Editorial Board, members of the Committee of the British Society for the Philosophy of Science and by an Advisory Panel which includes:

Peter Clark
Jon Dorling
Arthur Fine
Allan Franklin
Ian Hacking
Colin Howson
Noretta Koertge
Hugh Mellor
David Miller
Alan Musgrave
John Worrall
Elie Zahar

THE BRITISH SOCIETY FOR THE PHILOSOPHY OF SCIENCE

Committee 1984-85

President: Dr. J. Giedymin
President-Elect: Dr. D. H. Mellor
Hon. Secretary: Dr. P. Gelpke
Hon. Treasurer: Dr. K. V. Wilkes
Members: Mr. G. M. K. Hunt, Dr. W. H. Newton-Smith,
Professor H. R. Post, Mr. A. Young

The British Journal for the Philosophy of Science is the official organ of the Society, whose purpose is described by the following statement from its Constitution and Rules:

'The purpose of the Society is to study the logic, the method and the philosophy of science as well as those of the various special sciences, including the social sciences.'

The Society holds meetings, at which papers are read and discussed, once a month during the academic year. These are held, by kind permission, in University College, London. The Committee may recognise Regional Groups of Members formed to hold meetings in other parts of the country. Any Regional Group so recognised may nominate a representative Member to the Committee.

Ordinary membership is open to all interested persons on payment of the annual subscription. Membership of the Society is international and is by no means confined to people resident in the United Kingdom. The annual subscription is £4.00 for U.K. residents and £2.00 for overseas residents. Those who wish to join should write to the Hon. Secretary, Dr. P. Gelpke, Kingston College of Further Education, Kingston Hall Road, Kingston-upon-Thames, Surrey KT1 2TW.

Brit. J. Phil. Sci. Vol. 35, No. 2, June 1984

The British Journal for the Philosophy of Science

18 JUL 1985

Contents

Articles

101–113 J. BUTTERFIELD Relationism and Possible Worlds

115–128 EDWARD ERWIN The Standing of Psychoanalysis

129–147 BENNY SHANON Meno—a Cognitive Psychological View

Discussions

149–156 M. E. GRENANDER Behaviour Therapy: a Response to Edward Erwin

156–159 TYRONE LAI The Philosophical Relevance of 'Technically Good' Experiments

Review Articles

161–173 I. J. GOOD A Bayesian Approach in the Philosophy of Inference (review of PAUL HORWICH: *Probability and Evidence*)

HOMA KATOUZIAN T. S. Kuhn, Functionalism, and Sociology of Knowledge (review of B. BARNES: *T. S. Kuhn and Social Science*)

Reviews

175–193 D. H. Mellor on H. E. KYBURG: *Epistemology and Inference*; Gordon Blair on BRYAN MAGEE: *The Philosophy of Schopenhauer*; Robert Nola on GERARD RADNITSKY and GUNNAR ANDERSSON (*eds.*): *The Structure and Development of Science*; Derek Raine on N. T. ROSEVEARE: *Mercury's Perihelion from Le Verrier to Einstein*; Simon Schaffer on CHARLES WEBSTER: *From Paracelsus to Newton: magic and the making of modern science.*

Recent Publications

Announcement *Junior Postdoctoral Fellowships*

Printed in Great Britain at Aberdeen University Press

Volume 35 Number 3 September 1984

BJPIA 35 (No. 3) 201–311 (1984) ISSN 0007–0882

The British Journal for the Philosophy of Science

Published for the
British Society for the Philosophy of Science
by Aberdeen University Press

THE BRITISH JOURNAL FOR THE PHILOSOPHY OF SCIENCE

Editor: DR. DONALD A. GILLIES

Subscriptions

The *Journal* is published quarterly. The subscription price for mailing addresses in Britain is £14.00 net per volume post free and for mailing addresses outside Britain £17.00 (or foreign currency equivalent calculated at the day's rate); single issues £5.00 net (overseas £6.00) plus postage. Orders may be sent to any bookseller or direct to Aberdeen University Press, Farmers Hall, Aberdeen, Scotland AB9 2XT.

For members of the British Society for the Philosophy of Science—who should order direct from Aberdeen University Press—the subscription price is £9.00 (overseas £12.00). Membership of the Society is international and is by no means confined to people resident in the United Kingdom.

Microform

The current volume of the *Journal* is available in microform to subscribers to the printed version, *at a combined print/microfiche rate of* £19.50 (*overseas* £22.50).

Back Volumes

Volumes 1–16 may be obtained from Kraus Reprint Ltd., FL-9491 Nendeln, Liechtenstein or Kraus Reprint Corporation, 16 East 46 Street, New York, N.Y. 10017. Volumes 17–34 may be obtained from Aberdeen University Press, Farmers Hall, Aberdeen, Scotland AB9 2XT. Prices on application. Volumes 1–33 are also available in microform from Kraus Reprint Corporation.

Typescripts and Review Books

Typescripts, correspondence and books for review should be sent to the Editor at the Department of History and Philosophy of Science, Chelsea College, University of London, Manresa Road, London SW3 6LX.

Review books are accepted on the understanding that, although they will appear in the 'Recent Publications' list, a full review cannot be guaranteed, and that books not reviewed are not returnable.

Notes to Contributors

Three copies of all contributions must be sent. These need not, on initial submission, be in any particular style, provided they are typewritten and clearly legible. However, contributions accepted for publication may have to be wholly or partly retyped to accord with the *Journal*'s usual format—details of which may be obtained from the Editor. Contributions are considered on the assumption that they have not been published, wholly or in part, elsewhere.

Articles should be accompanied by abstracts of about 100 words.

Material is accepted for publication in the *Journal* on the condition that copyright is assigned to the British Society for the Philosophy of Science.

© British Society for the Philosophy of Science 1984

For permission to reproduce material from *The British Journal for the Philosophy of Science,* please apply to the Editor.

Brit. J. Phil. Sci. 35 (1984), 201–211 *Printed in Great Britain* 201

Truth, Rationality and the Sociology of Science

by RICHARD C. JENNINGS*

Philosophers of science are becoming more sensitive to the claims about truth and rationality being made by sociologists of science. There is a tendency among some of these philosophers to dismiss such claims as irrelevant to philosophy of science and as self-refuting. Larry Laudan, in his 'arationality assumption', has captured the essence of positions which argue that sociology of science can only be concerned with scientific claims which are not rational (or, in some versions, 'not true'). I show that the arguments he gives in support of the arationality assumption are not sound, that the arationality assumption has absurd consequences, and I argue that the arationality assumption undermines the possibility of fruitful cooperation between philosophers and sociologists of science. I conclude that the arationality assumption should be rejected.

1 *Failure of Arguments Supporting the Arationality Assumption*
2 *Absurd Consequences of the Arationality Assumption*
3 *Pernicious Consequences of the Arationality Assumption*
4 *Conclusion*

The recent turn in philosophy of science toward historical studies of science has been accompanied by a growing interest in the sociology of science. In the last chapter of *Progress and its Problems* Larry Laudan [1977][1] has focused his attention on what he calls cognitive sociology. In contrast to a sociology which studies, say, the societies and institutions of science or the changing numbers of students opting for science courses, cognitive sociology of science seeks to explain the conceptual content of theories, their discovery and their acceptance or rejection in terms of social or economic factors. Laudan characterises cognitive sociology of science as follows:

> . . . a sociologist may seek to explain why a certain *theory* was discovered (or, after discovery, accepted or rejected) by pointing to the social or economic factors that predisposed scientists to be sympathetic or hostile to it. Alternatively, he may seek to show that certain social structures were influential on the genesis of the concepts of a theory. Such efforts fall within the scope of what I shall call the cognitive sociology of Science (p. 197).

Laudan contrasts cognitive sociology with 'the intellectual or rational historiography of science' (p. 98). Here he finds '. . . the possibility for enormous (and potentially fruitful) conflict' (p. 198). I differ from Laudan in finding between cognitive sociology and the intellectual or rational historiography of science the possibility for enormous and potentially fruitful cooperation. But before the fruits of this cooperation can be realised, its possibility must be established, and this is the aim of the present paper.

---

* A shortened version of this paper was read in Boston, Massachusetts, at the Eastern Division meeting of the American Philosophical Association during December 1983. I would like to thank my commentator, Lowell Kleiman, the chairman of the session, Harvey Siegel, and the members of the audience for an enjoyable and clarifying discussion.

[1] In the following discussion, passages from Laudan's [1977] will be cited by enclosing the page number of the citation in parentheses at the end of the passage.

The main obstacle to such cooperation is the belief that the domain of cognitive sociology is incompatible with that of philosophy of science. This is the belief, stated in various ways, that cognitive sociologists should occupy themselves with the study of false beliefs or irrational decisions while the philosophers of science should attend to the establishment of true beliefs and the making of rational decisions. This belief in the asymmetry between cognitive sociology and philosophy of science is formulated as an assumption by Laudan—he calls it the arationality assumption. The arationality assumption is:

> . . . *the claim that the sociology of knowledge may step in to explain beliefs if and only if those beliefs cannot be explained in terms of their rational merits* (p. 202, Laudan's italics).

He goes on to say that:

> Essentially, the arationality assumption establishes a division of labor between the historian of ideas and the sociologist of knowledge . . . (p. 202).

This second passage and further passages make it abundantly clear that one of Laudan's main preoccupations (whether conscious or not) is to maintain the boundaries between sociology of science and philosophy of science (or rational historiography of science). I believe, and will argue, that it is counter-productive to attempt to maintain such a boundary and that there are difficulties facing attempts to define such a boundary—in particular there are serious difficulties in maintaining the arationality assumption. Now, in fact, this assumption is explicitly rejected by many contemporary cognitive sociologists, or sociologists of science: it is far from being universally embraced and much fruitful study now simply ignores this assumption. When I refer in this paper to contemporary cognitive sociologists, or sociologists of science, I have in mind mainly those of the University of Edinburgh Science Study Unit—Barry Barnes, David Bloor, David Edge and Steven Shapin. In Bath, Harry Collins and Trevor Pinch are working along these same lines as are Karin D. Knorr-Cetina and Thomas Nickles in the USA. For some programmatic statements and characteristic work of these contemporary cognitive sociologists I suggest the following: Barnes [1974], Bloor [1976], Barnes and Edge [1982] and Shapin [1982]. Laudan's remark that:

> . . . most cognitive sociologists adhere to the arationality assumption . . . [and that] whatever its grounds might be, it is taken as axiomatic by most practitioners . . . (p. 205).

is now no longer true, though in the early seventies that may well have been the way things looked. I am therefore concerned with Laudan's justifications for accepting the arationality assumption and I will argue not only that he fails to justify its acceptance but that the assumption itself is subject to a *reductio ad absurdum* argument, and that it has pernicious effects on the possibility of cooperation between philosophers and sociologists of science.

I FAILURE OF ARGUMENTS SUPPORTING THE ARATIONALITY ASSUMPTION

The first argument that Laudan offers in support of the arationality assumption is the now familiar argument from self-reference. Laudan presents the argument as follows:

> If it were true that *all* beliefs were not the result of rational deliberation or enlightened evaluation, but rather were simply determined by the social situation of the believer, then the whole enterprise of cognitive sociology would be self-indicting; for if *all* beliefs are socially caused, rather than rationally well-founded, then the beliefs of the cognitive sociologist himself have no relevant rational credentials and hence no special claims to acceptability. Ernst Grünwald put the point tellingly when he remarked: 'For the thesis that all thought is existentially [i.e., socially] determined and thus cannot claim to be true does itself claim to be true.' Thus, the cognitive sociologist, to avoid being hoisted with his own petard, is committed to the view that *some* beliefs are rationally well-founded rather than socially determined (p. 201).

But if this argument is to have the force that Laudan (and others) take it to have, then it is necessary to employ the very arationality assumption that the argument is meant to support. Let us see how this comes about.

The argument begins by correctly pointing out that if all knowledge or belief is to be explained socially, then those very explanations themselves can be explained socially. Again, and in particular, if the evolution of scientific thought is to be explained by cognitive sociology then cognitive sociology itself will be subject to explanation by cognitive sociology. Cognitive sociology will find itself explaining its own cognitive processes. As this argument stands at this point there is no obvious problem. To be sure, cognitive sociology is self-referential, and there have been problems with self-reference in the past. But there is no obvious problem here. Indeed, contemporary sociologists of science have accepted the reflexive nature of their study and are all too happy to indulge in the auto-analytic endeavour of discovering the sources of their own views on the sources of scientific knowledge.

The problem in the self-reference of cognitive sociology only arises if a further assumption is made to the effect that beliefs which are, or can be, explained sociologically are in some way defective—whether false, prejudiced, irrational or whatever. And here we need the arationality assumption itself: the assumption that sociology of knowledge may step in to explain beliefs (if and) *only if* those beliefs cannot be explained in terms of their rational merits. In Laudan's discussion the fault is that of irrationality although he calls in the support of Ernst Grünwald who says that socially determined thought cannot claim to be true. The distinction between claims that are acceptable and those that are defective is crucial to this passage—the final remark, that the cognitive sociologist is trying to explode the claims of others through his analysis, is an extreme use of the distinction and completely contrary to the practice of contemporary sociologists of science. If we agree that beliefs which can be explained sociologically are defective, that is, false or irrational, then the argument from self-reference established that the claims of cognitive sociology are defective—cognitive sociology turns out to be self-defeating. But if we do *not* assume that the provision of a sociological explanation of knowledge implies that the knowledge is defective, then the fact that the cognitive sociologist must also take into account his own knowledge claims does not imply that those knowledge claims are defective, and cognitive sociology does not turn out to be self-defeating. The arationality assumption is needed to argue from the correct claim that cognitive sociology is self-referential to the incorrect claim that it is irrational or false. Self-reference is not a theoretical reason for placing limits on cognitive sociology—cognitive sociology does not have to be protected from itself. The first argument in support of the arationality assumption, the argument from self-reference, depends on the arationality assumption itself as one of the premisses. This argument therefore fails through being circular.

The second argument that Laudan offers in support of the arationality assumption is based on an imaginary example. We are asked to imagine two theorists, *y* and *z*, who give explanations to a belief, *A*, held by someone else, say *x*. According to the story, *y* explains *x*'s belief in *A* in terms of its being well founded on other beliefs held by *x*, while *z* offers an explanation based on the social/psychological background of *x*. Laudan then asks, 'How, if at all, can *y* convince *z* that his explanation is bogus because it violates the arationality assumption?' (p. 203). The way to do this he suggests (if we are not to simply take the arationality assumption on faith), is to point out to *z* that in his own explanation of *x*'s belief in *A*, he is offering an argument, a reasoning process which moves from his own premisses to his conclusion. If *z* claims that his own conclusions are rationally well founded and not merely a function of his social position, then, Laudan argues, the onus is on him to show why his own beliefs are 'situationally transcendent' (p. 204) while that of *x* is psycho-socially determined—especially given that *y* can provide an explanation of *x*'s belief in terms of its being rationally well founded on other of *x*'s beliefs. This argument again rests on the arationality assumption, the assumption that if a sociological explanation of certain beliefs can be given then those beliefs *cannot* be accounted for rationally. If there were no stigma attached to beliefs which were explained sociologically, then there would be no onus for *z* to give a 'situationally transcendent' account of his own beliefs. In the course of this argument Laudan makes the parenthetical remark that:

> One presumes here that *z* would not take kindly to the suggestion that the only cause of his belief in a certain sociological explanation was his location in the social network! (p. 204)

But of course this is an elementary matter of self-consciousness with which contemporary cognitive sociologists of science are fully conversant. This second argument, like the first one, depends on the arationality assumption and therefore fails through being circular.

In this second argument, however, one of my main concerns comes to the surface, namely, the suggested, but non-existent, conflict of interest between the tasks that *y* and *z* have undertaken. The example that Laudan has formulated is over-simplified and presented *as* a case of conflict between *y* who gives a rational account of *x*'s belief in *A* and *z* who gives a social/psychological account of *x*'s belief in *A*. What is missing from the imaginary story is any account of how *z* explains the process and mechanism of *x*'s coming to accept the belief in *A*. The story that the cognitive sociologist tells is not a story of a black-box into which various psychological and sociological forces play, and out of which ultimately comes a belief; rather he tells how these forces condition the other beliefs that *x* holds, how connections between various beliefs are strengthened or weakened as a result of acquiring new beliefs and interests. The structure of *x*'s beliefs is most likely a rational structure, fully compatible with *y*'s analysis. *Z*'s analysis is an analysis of how *that* structure came to be, rather than some other possible structure. In a word, I do not see the tasks of *y* and *z* as incompatible or conflicting, but as complementary and mutually supporting.

Laudan takes it as virtually axiomatic that these two approaches to explanation of belief will be in conflict. In the introduction to his discussion of cognitive sociology he claims that, '. . . here, there is the possibility for enormous (and potentially fruitful) conflict' (p. 198). He goes on to argue that:

> The intellectual historian of knowledge will generally seek to explain why some agent believed some theory by talking about the arguments and the evidence for and against the theory and its competitors. The cognitive sociologist of knowledge, on the other hand, will generally try to explain why the agent believed the theory in terms of the social economic, psychological, and institutional circumstances in which the agent found himself. Both are trying to solve the same problem (namely, the belief of some historical agent), yet their modes of solution are so different as to be almost incommensurable. Is there any way, given these conflicting explanatory strategies, of determining who is right, the intellectual historian or the cognitive sociologist? Or could they both be? (p. 198)

I believe that they *can* both be right, and that there is nothing paradoxical about this. I believe that they are looking at the same processes from different points of view and can offer complementary and compatible accounts of how these processes take place. Roughly, the ground of this claim consists in the recognition that there is no hard and fast account of rationality, true for all times and places, and that what is taken to be rationally justified is as much subject to what is wanted to be justified as to developing theories of rationality.

The third argument for accepting the arationality assumption is given in terms again of the conflict between *y* and *z*. According to this argument we should consider the theoretical systems of *y* and *z* as competing research traditions. Then, the argument continues, we can see that the rational historiography of ideas has gone much further than historical sociology has toward explaining historical cases in the evolution of science. 'Indeed,' Laudan claims, 'the "success ratio" of intellectual history is several orders of magnitude greater than that of cognitive sociology' (p. 204). In this event, he concludes, good sense dictates that rational explanations should be given priority over sociological ones since they have been the most successful. This argument is, of course, the argument inevitably used in the preservation of any *status quo*, namely, that the method has worked up to now so let us keep using it. However, in considering the type of explanation offered by *z* as an instance of a research tradition, Laudan has come remarkably close to the attitude taken by contemporary sociologists of science. For them, these sorts of explanations are being tried out much as a hypothesis in science is tried out—contemporary sociologists of science are looking at matters case-by-case to see how well their kind of analysis works. For them it *is* an empirical question when and how well their kind of analysis works. And one of the essential features of the contemporary programme is the rejection of the arationality assumption. The fact (if it still is a fact) that not so much has been explained yet by cognitive sociology as by rational historiography is not so exciting for the contemporary cognitive sociologist as the remarkable success that he has had and is having in the actual applications of his programme. For a survey of the success of contemporary sociology of science, see Shapin [1982] and the extensive analytical bibliography contained therein. If we were to consider the excitement generated by a research tradition or its rate of growth, then cognitive sociology would, in Laudan's terms, be given clear priority over rational historiography. The third argument is at best inconclusive and, if pushed, more likely to turn against Laudan than to help him. Moreover, we once again find the arationality assumption being employed to create the illusion that there is a conflict between doing cognitive sociology and rational historiography.

Believing himself to have established the arationality assumption through the three arguments discussed above, Laudan goes on to criticise cognitive sociologists for their concepts of rationality. I will not consider this immediately but instead first attend to what Laudan sees as three important consequences of accepting the arationality assumption.

2 ABSURD CONSEQUENCES OF THE ARATIONALITY ASSUMPTION

Laudan draws three important consequences from the arationality assumption. Each of these consequences shares with the arationality assumption the essential prerequisite that we be able to determine whether a belief is held rationally or not. And in their requirement that we be able to show the rationality of beliefs, these three consequences clearly exhibit the difficulties to be found in working with the arationality assumption. What then are these important consequences? Let me quote Laudan:

> We thus see that accepting the arationality assumption has three important consequences: (1) the domain of possible belief situations for sociological analysis is restricted to those in which agents accept beliefs or weight problems in ways incompatible with what rational appraisals would suggest; (2) the sociologist of knowledge must be able to show that the theory of rationality to which he subscribes (in order to determine what cases might be sociological) is the best available; (3) the historical sociologist of knowledge must show, for any given episode he wishes to explain, that it is incapable of being explained in terms of rational, intellectual history (p. 209).

The net effect of these consequences is to completely foreclose the possibility of doing cognitive sociology and, given that Laudan seems to have some (albeit remarkably limited) sympathy with cognitive sociology, I take this to be a ***reductio ad absurdum*** of accepting the arationality assumption. The first consequence is not particularly problematic, it merely serves to limit the field of application of cognitive sociology. But the second consequence raises a very difficult problem, one that even Laudan does not boldly claim to have solved, namely the problem of showing that some theory of rationality or other is the best available one. The theory of ***Progress and its Problems*** may be Laudan's choice for the best available theory but to actually *show* that it is the best is quite another matter. To show that some theory of rationality is the best available one is a problem which philosophers of science, whose business it is, have not solved. Yet what Laudan claims in drawing this second consequence is that the cognitive sociologist must show that some theory of rationality is the best available one in order to get on with the task of discovering which cases of theory choice he can hope to analyse (namely those which *do not* fit the theory he has shown to be the best). The third consequence requires a notoriously difficult kind of proof, namely a non-existence proof. The cognitive sociologist is obliged to show for any given episode he wishes to explain that ***there is no explanation*** of that episode in terms of rational, intellectual history. Thus it appears that the cognitive sociologist is obliged to work through every conceivable explanation in terms of rational, intellectual history and show them all to be unsuccessful before he can offer his own explanation. The impracticability of this, if not its outright impossibility, should be patently obvious. It is apparent that if the cognitive sociologist is to take these consequences seriously and act on them he will never be able to get started doing his cognitive sociology.

How Laudan can so easily draw consequences from his arationality assumption that impose such an obvious burden on the cognitive sociologist with no apparent recognition of this fact is difficult to understand. A possible analysis of what Laudan is up to here is available in terms of cognitive sociology—indeed, an analysis which may not violate the arationality assumption! Laudan, (so the analysis goes) being a philosopher of science, believes that certain sorts of questions are within his domain of competence. He is sensitive to having any of his domain encroached upon and is therefore highly critical of any other solutions to the sort of problems that he regards as his own. He therefore makes as a prerequisite for access to those problems the fulfilment of impossible tasks. And one of the tasks is to show beyond a doubt that the problems are not ones in which Laudan will ever take an interest. The basic explanatory concept operating here is not very sophisticated—it is simply the idea of academic territoriality, or professional boundary maintenance.

3 PERNICIOUS CONSEQUENCES OF THE ARATIONALITY ASSUMPTION

Having shown that Laudan's arguments for the arationality assumption do not actually work, and that the assumption itself has absurd consequences, I will now discuss a more serious problem with the assumption. The problem that I see in adopting the arationality assumption is that it leads to divisiveness where there need be none, and to the disruption of a potentially viable programme of cooperative work between cognitive sociologists and philosophers of science. One of the main contributions that Laudan sees himself making to the theory of scientific rationality is in the broadening of that notion to include the solutions to conceptual problems and even 'ideological integration' (p. 214). He claims that such factors as the moral, religious, epistemological and metaphysical consequences of theories are relevant to the rational analysis of scientific decision making. Thus, consider the following claim summarising a passage in which he argues the need to broaden our concept of rationality:

> I am simply suggesting that we need a broadened notion of rationality which will show how the 'intrusion' of seemingly 'nonscientific' factors into scientific decision making is, or can be, an entirely rational process. Far from viewing the introduction of philosophical, religious and moral issues into science as the triumph of prejudice, superstition and irrationality, this model claims that the presence of such elements may be entirely rational; further, that the suppression of such elements may itself be irrational and prejudicial (p. 132).

Or consider this claim:

> Thus, contrary to common belief, it can be rational to raise philosophical and religious objections against a particular theory or research tradition, if the latter runs

> counter to a well-established part of our general *Weltbild*—even if that *Weltbild* is not 'scientific' (in the usual sense of the word) (p. 124).

Or, finally:

> We must not assume *a priori*, as some historians of science have, that the only important parameters were experimental or other obviously 'scientific' ones. Because theories and research traditions have to be accommodated within a broader network of beliefs and preconceptions, any accurate appraisal of an episode must attend carefully to the philosophical, theological and other intellectual currents which were brought to bear on the case at hand (p. 128).

The problem created here by the arationality assumption is that it implies that if these factors are included in rational analyses of theory choice they can *not* be included in the analyses of cognitive sociology.

Apart from the problem of deciding when such factors do or do not belong to a rational story, the idea that such a decision *must be made* creates an unnecessary division between the tasks of the sociologist and the philosopher of science. An immediate manifestation of this problem is seen in Laudan's discussion of sociologists of science. Laudan adopts a surprisingly arrogant and chauvinistic stance with respect to cognitive sociologists. His arrogance consists mainly in criticising their concepts of rationality. He criticises them for '. . . their failure to recognise that there may be a broad spectrum of theories of rationality . . .' (p. 206) and for their tendency to choose '. . . that model of rationality which is the most limited of all' (p. 206). Of T. S. Kuhn he says:

> . . . one wishes he might have worried more deeply about what rationality amounts to before jumping to the conclusion that his empirical model of rationality was subtile enough to provide the careful discriminations between the immanent and the arational (p. 207).

He concludes his arrogant remarks with the claim that he has:

> . . . dwelt at length on the parasitic dependence of cognitive sociology of knowledge on theories of rationality not only to draw attention to the need for sociologists to be more self-critical about their judgments concerning the rationality of particular cases, but [and here the chauvinism begins] also to emphasise the fact that *the application of cognitive sociology to historical cases must await the prior results of the application of the methods of intellectual history to those cases* (p. 208, Laudan's italics).

A further example of his chauvinism is in the remark that:

> . . . the cognitive sociologist must simply bide his time; to do otherwise is to abrogate the arationality assumption which is at the heart of contemporary sociological thought (p. 209).

This latter remark may simply show that Laudan's book is out of date in its assessment of what sociologists of science do. But Laudan does seem to be aware of current trends in the sociology of science from the parenthetical remark immediately following the one just quoted:

> Something akin to this point was recognized by Mannheim, but his latterday disciples have tended to assume that one could do sociological history in blissful ignorance of the rational history of ideas (p. 209).

I take these 'latterday disciples' to be contemporary sociologists of science, but they are not 'in blissful ignorance of the rational history of science', rather, they self-consciously reject the arationality assumption in order to give a full and complete account of how scientific ideas evolve—very much in the way Laudan seems to want to do, except that they are not concerned to try and distinguish that which should count as rational from that which should not.

This arrogance and chauvinism is surprising in light of the fact that cognitive sociologists have been studying the very factors that Laudan claims are relevant to his own analysis of science. Instead of welcoming their work, Laudan criticizes the cognitive sociologists who have done such work for doing work which falls *outside* their realm! This criticism appears in the following remark:

> What should be pointed out here is that there are whole 'schools' of cognitive sociology (one thinks especially of Sorokin, Scheler, Durkheim and Richter, for example) which see the central aim of sociology to be a study of the ways in which the different ideological elements in a culture are integrated. If the arguments of this essay have any cogency, studies of 'ideological integration,' in so far as such integration is rationally well-founded, belong to intellectual history and fall completely *outside* the realm of cognitive sociology (p. 214).

In view of the similarity of his interests and those of cognitive sociologists it is not at all clear why Laudan insists on using the arationality assumption to drive a wedge between two closely allied and potentially cooperative fields. Or, if it is clear, it is rather depressing. The only clear explanation I can give for this sort of criticism of cognitive sociology is that Laudan is engaged in a kind of professional imperialism—he is attempting to colonise an adjacent field of academic work by broadening the scope of his own field and denying (by means of the arationality assumption) that two professions can cooperate in working the field.

A further consequence of accepting the arationality assumption is that if we do decide some choices are not amenable to rational analysis we are in the position of castigating some historical actor (or actors) as irrationally subject to social forces. But in the ordinary course of scientific development individuals make choices which are supported by reasons and justifications. To simply reject the reasons and justifications of some historical actor is to impose external standards of rationality which do not fit the actor's own standards—we are, in a word, denying the actor's own rationality. Now it is certainly true that there is no universally recognised and adhered to analysis of rationality, and no single algorithmic way of making a rational choice. There are a variety of factors that can be taken into consideration in any decision, and what sorts of arguments the actor uses and what factors he takes into consideration are his own choice. The story that the actor has to tell about his choice is a story of rational choice. If we want to distinguish the task of the philosopher of science from that of the cognitive sociologist, we could say that the philosopher of science studies the logic or rationality of the story that the actor tells, taking into account all of the various factors that the actor considers relevant, while the cognitive sociologist studies how the actor came to construct the particular rational story that he did from the various sources available to him. Thus when Laudan claims of the two examples of cognitive sociology that he actually looks at (to each of which he devotes a whole paragraph) that neither of them consider the possibility that the choices they study '. . . might have been an entirely appropriate and rational response to the empirical and conceptual criticisms of previously dominant theories' (p. 216), he misses the point. There is no question that they *were* appropriate rational responses to the previous theories, the question is why *those* particular rational responses were exhibited, rather than some *other* rational responses. Why, in other words, did the actors under discussion choose to construct one rational story rather than another? And it is the answer to *this* question that the cognitive sociologist is concerned to provide.

4 CONCLUSION

Having shown that the three arguments Laudan uses to support the arationality assumption do not work, that three important consequences of the arationality assumption have absurd consequences, and that the arationality assumption has pernicious consequences for the study of the history of science, I conclude that we should reject the arationality assumption.

University of Cambridge

REFERENCES

BARNES, B. [1974]: *Scientific Knowledge and Sociological Theory*. Routledge and Kegan Paul.
BARNES, B. and EDGE, D. (*eds*.) [1976]: *Science in Context*. Open University Press.
BLOOR, D. [1976]: *Knowledge and Social Imagery*. Routledge and Kegan Paul.
LAUDAN, L. [1977]: *Progress and its Problems*. Routledge and Kegan Paul.
SHAPIN, S. [1982]: 'History of Science and its Sociological Reconstructions', *History of Science*, **20**, pp. 157–211.

Brit. J. Phil. Sci. 35 (1984), 213–221 *Printed in Great Britain*

The Calibration Question

by FRANK LAD*

Recent discussion of the calibration of probability assessments is related to the earlier influential attitudes of Fréchet. The limiting frequency criterion of good calibration is criticised as being of no relevance to the evaluation of the probability of any event. An operational definition of good calibration is proposed which treats calibration properties as characteristics of the assessor's entire body of opinion, not of opinion about some particular event or events. In these terms a result is shown which says that every coherent opinion distribution is well-calibrated. Scoring rule representations as calibration plus refinement scores are discussed. It is proposed that a subjectivist could well attribute particular bad experiences to 'luck', and must balance this sense of luck with a sense of the inadequacy of his/her own knowledge when evaluating previous knowledge and experience.

1 *Introduction*
2 *A Critique of the Limiting Frequency Criterion of Calibration*
3 *A Reconstruction of the Calibration Issue*
4 *An Inherent Brier Score Equivalency of All Coherent Distributions*
5 *On Luck and Incoherence*

1 INTRODUCTION

Calibration characteristics of personal probability assessments have received considerable attention over the past decade. As commonly defined, a person's probability assessments are considered to be well-calibrated at value p if the limiting frequency of occurrence of events which are assessed with probability p, equals p. What might be called 'the calibration question' has been succinctly specified by Pratt [1962]: 'Must subjective probabilities be realized by relative frequencies?' Positive responses have been usually motivated by the Bernoulli law of large numbers.

Analysis has focused on the relevance of this calibration criterion to more general theoretical problems of scoring a probability distribution, and on the empirical verification of the calibration status of particular forecasters, notably weather forecasters and subjects of psychological experimentation. Lichtenstein, Fischoff and Phillips [1982] have reviewed empirical pro-

Received 3 July 1983

* Thanks to Phil Dawid, Morris DeGroot, Lance Girton, Sarah Lichtenstein, David Lane and John Pratt for helpful discussion and communication. The views presented are my own. An earlier version of this paper was presented to the 24th meeting of the NBER-NSF Seminar on Bayesian Inference in Econometrics, 8 May 1982 at the State University of New York at Albany.

cedures and results reported through 1980. DeGroot and Fienberg [1983] discuss the relation of calibration to more general theoretical analyses of scoring rules. Recently questions have focused on the possibility that the assessments of a coherent probability assessor may in fact be uncalibrated according to the definition above (Dawid [1982], and Kadane and Lichtenstein [1982]). Overall, these discussions flirt with the 'half baked ideas' raised by de Finetti [1962] in asking 'Does It Make Sense to Speak of "Good Probability Appraisers"?'

Interestingly, the calibration question was a controversial one at the International Congress on Philosophy of Science in Paris, 1949, as evidenced by Fréchet's report [1949] with supplementary comments by de Finetti. Fréchet took the calibration criterion so seriously that it had led him to conclude that only subjective probabilities which are equivalent to frequencies can be regarded as valid.

> Suppose that a degree of belief is small for events which are not rare. If someone acts upon that real degree of belief as the base of his conduct, we could observe the considered individual out walking in his shirt sleeves in the middle of winter and all bundled up in summer, or promenading tranquilly along the guardrail of a bridge, *etc.* because he finds himself to have attributed a small degree of belief to the possibility that it is cold in winter, hot in summer or that one falls when climbing the guardrail. If people would act this way they would disappear rapidly from the face of the earth and with them this unfortunate [subjective] interpretation of probability ... either the degree of belief is estimated badly enough to lead to their death those who made it the foundation of their conduct, or else the degree of belief is not at all subjective and is estimated approximately by a frequency [1946, pp. 199–201].

Frechet's specification of long run calibration as the criterion for the correctness of a probability assessment was explicit at the Congress.

> ... But how would one judge if his estimations of probability are sufficiently correct? Would it be when an event of which he judges the probability greater than 1/2 does not occur? No, certainly not if this circumstance occurs only one time. It only incites him to revise his estimation if the frequency of this event remains small in a great enough number of trials.
>
> Thus this frequency which our subjectivist would like to ignore returns to him like a boomerang to astonish him and lead him to repentance [1949, pp. 220–1].

De Finetti was disturbed by these arguments and other reactions to subjective views at the 1949 Congress, as he remarks in the beginning of his *Theory of Probability* [1974, p. 1 and p. xiv]. Aware of their fundamental difference, Fréchet invited de Finetti to append his misgivings to the report. This addendum is included in the collection of Fréchet [1955a] to which page number references for the quotations in this paper refer.

De Finetti's views on the issue are contained in his extensive discussion of 'Frequency and "Wisdom After the Event"' [1974, section 5.9]. His argument is based on the fact that the probability assessments specified for a sequence of quantities represent the state of uncertainty of the assessor given his/her state of information. Since probabilities (previsions) are not predictions, there is no point in comparing the previsions with the outcomes as if it made sense to ask whether they had been correct or not. Posterior probabilities of subsequent events are not corrections of previous probabilities. They are simply a new evaluation, cohering with the previous one, corresponding to the new state of information.

This view is in direct contrast to Fréchet's statement that upon observing a long string of outcomes for which the observed frequency were far from their assessed probability, 'I would not be long in attributing no value to those degrees of belief which I had previously evaluated' [1945, p. 177].

Subsequent literature has mainly pursued the calibration question from a standpoint in accord with Fréchet's attitude as opposed to de Finetti's. This is apparent in the general acceptance of the limiting frequency specification of the calibration criterion.

The following argument develops a new analysis of the calibration issue wholly consistent with the de Finetti outlook. I shall first criticise the limiting frequency criterion as being irrelevant to any sense of the 'correctness' of a probability assessment. Then, according to a more expansive definition of calibration, I shall argue that good calibration is equivalent to coherence. Examination of the so-called Brier score (calibration plus refinement) of a sequence of probability assessments motivates a specification of the meaning of luck within the operational subjective viewpoint.

The various points of view on this question are relevant to contemporary theoretical work in economics, particularly to the 'rational expectations' hypothesis *vis-à-vis* its assumption of 'correct' probabilities assessed by economic agents. (See, for example, Lucas [1973] and the collection of Lucas and Sargent [1981].) In this regard it is notable that Fréchet himself, who demanded good calibration as a characteristic of correct probability assessments, insisted (Fréchet [1955b]) that econometric analysis recognise the prevalence of irrational expectations motivating actual human behaviour.

2 A CRITIQUE OF THE LIMITING FREQUENCY CRITERION OF CALIBRATION

Given the extreme conclusions drawn by Fréchet from considerations of probability calibration based on frequencies, it is important to clarify precisely which object can be calibrated by such a scheme. Is it some event in question, the probability of that event in question, or the collection of all probability evaluations made by some particular appraiser? I believe it is only the last of these alternatives which can be meaningfully analysed.

One should notice that the good calibration of an appraiser according to the relative frequency criterion cannot support the assertion that any particular prevision (assessed probability) appraised by that assessor is 'correct'. Ignoring the operational difficulty associated with observing a limiting frequency, let us suppose we have before us a person known to be well-calibrated as defined. If this assessor were to announce a probability for some specified event as p, would the assessor's characteristic of being well-calibrated give others any reason to conclude that the correct probability of that event equals p? No, because in regards to that event, no matter what value of p is announced and no matter whether the event occurs or not, this person will still be regarded as well-calibrated according to the definition. Appending a single zero or one to any infinite string of indicators for which the proportion of ones equals p will not change that proportion for the 'enlarged' string. Moreover, two appraisers, each well-calibrated according to this definition, could easily assess their probabilities of the same event at different values. For the appraisers' good calibration does not characterise the event, but their probability assessments of that event. Each appraiser has a personal choice of the value at which to assess p for any particular event.

The only string of probability assessments which cannot be criticised on any ground is that string which associates probability one with the observed outcomes of history and probability zero with outcomes which do not occur. It is because no one knows precisely which of all possible degenerate distributions is this 'correct' one that we use full distributions to express our knowledge. When person A defers to person B for making a decision, it is not because A feels that B is a better estimator of some correct unknown probability of some event, but because person A allows that B knows more about the event in question.

The calibration quality of a probability assessor would better be considered not as a characteristic of judgements about any particular event or sequence of events, but as a characteristic of all probability assessments specified by that person. Notice that, similarly, the coherence of a probability distribution is judged not by the specified probability for some particular events, but by the composite of all probabilities specified. In the next section we find that good calibration, formally defined as a global characteristic of all specified assessments, is in fact implied by the requirement of coherence!

3 A RECONSTRUCTION OF THE CALIBRATION ISSUE

A different approach to calibration from that taken by calibrationists is to consider calibration quality a global characteristic of *all* the bets that a person is avowedly willing to make about each particular observable quantity. The following alternative operational definition is conceived in this spirit. It is proved subsequently that any coherent assessments specifying the n-tiles of a probability distribution are necessarily well-calibrated.

DEFINITION Probability assessments for the random quantity X are said to be well-calibrated at probability p if the observed proportion of *all* propositions regarding X assessed with probability p equals p.

THEOREM Suppose coherent assessments are elicited from an assessor who is to observe a random quantity X, and who has introspected sufficiently to specify the n n-tiles of his/her probability distribution, $X_{(1)} < X_{(2)} < \ldots < X_{(n)} = X_{\max}$. That is, the assessor's $P[X \leqslant X_{(j)}] = j/n$. Presume $X_{\min} \leqslant X_{(1)}$. Then no matter what the measured value of X, the assessed probabilities are well-calibrated at each $p = j/n$ for $j = 1, \ldots, n$.

PROOF For each value of j, the elicited n-tiles and the condition of coherence specify ${}_nC_j$ distinct proposition conjunctions with assessed probability j/n. Now the observed value of X will surely lie within one and only one of the n intervals defined by these n-tiles. Of those proposition conjunctions specified as entertained with probability j/n, the ones which in fact obtain are all those which include the fractile containing the observation value. The number of these equals ${}_{n-1}C_{j-1}$. Thus the proportion of all propositions with assessed probability j/n which in fact obtain equals ${}_{n-1}C_{j-1}/{}_nC_j = j/n$. Q.E.D.

COMMENTS The implication of the result above is more far reaching than a facevalue reading of the conditions suggests. Suppose, for example, that X is a Bernoulli random quantity with possible values of only zero or one. How might the above theorem demanding n assessed n-tiles, apply if one's prevision is some $p \neq \frac{1}{2}$, say $p = 0.8$? To the experiment X might be appended any auxiliary experiment toward which one can specify the four quartiles of one's opinion distribution, for example the four ordered outcomes of two flips of a coin in my pocket. If the outcome of the original random quantity $X = 1$ were partioned into $\{X = 1\} \equiv \{(X = 1 \text{ and } HH), (X = 1 \text{ and } HT), (X = 1 \text{ and } TH), (X = 1 \text{ and } TT)\}$, the distinct outcomes of that partition along with the outcome $X = 0$ would constitute a random quantity about which five quintiles of opinion are specified. The theorem above insures that exactly one-fifth of them shall be instantiated (and that k-fifths of all unions of k of them would be as well.) For any assessed prevision $p \neq \frac{1}{2}$, such an auxiliary experiment could be constructed. The argument mimics that of Savage [1954, section 3.3].

The statement and discussion of the theorem are couched in terms consistent with de Finetti's consideration that measurable quantities exhibit only a finite number of readings. The elicitation of one's subjective n-tiles is a common way of specifying a personal probability distribution relative to a limited amount of introspection. The gist of the result applies however to an appropriate extension of the setup to any fully specified continuous distribution for a random variable measurable with respect to the standard Borel sets of the real line.

The subjective definition of the probability of an event A as p implies that the expected gain from placing a bet on A for stake S at price pS equals zero. In fact the gain from placing such a bet will be either $S-pS$ if A is the case, or $-pS$ if not. The result just proved states further that if one placed *every* bet with which he/she associates probability p for stake S at price pS then the actual accumulated gain from all those bets equals zero.

Corollary to the above theorem is the recognition that incoherent assessments are not well-calibrated according to the definition proposed. For incoherence in some way amounts to the assessment of the probability of the sure event at some value other than one, while the frequency of occurence of that event rests at one. Incoherence involves a specification of too many events assessed with high probabilities and too few with low probabilities (or vice versa). Observed non-calibration according to the standard frequency definition for a large number of independent events might be explained by referral to the outcomes of other assessed events which were not scored. But it might also be taken as evidence that one's full corpus of previsions is incoherent in some unelicited way. We shall return to this notion later in a consideration of luck.

Further corollary to the theorem is the awareness that relative to *any* coherent n-tile appraisal there do exist particular sequences of indicators for which the observed frequency of occurence is far from the assessed probability. Consideration of such an apparent anomaly was sufficient for Fréchet to reject the probability assessments as without value. In the argument of this section I claim to have shown that it is only over the sequence of all assessed indicators that good calibration is desirable, and is implied by coherence.

4 AN INHERENT BRIER SCORE EQUIVALENCY OF ALL COHERENT DISTRIBUTIONS

The theorem implies further an interesting equivalence characteristic of every coherent specification of n-tiles for the random quantity X. It can be exposed through use of scoring rules for evaluating different probability assessments.

The well known Brier score [1950] of a sequence of previsions $P_1, \ldots, P_N$ corresponding to the indicators $I_1, \ldots, I_N$ is $BS_N = \sum_{i=1}^{N} N^{-1} (P_i - I_i)^2$. This same quadratic scoring rule is fundamental to de Finetti's definition of prevision [1974]. The work of Sanders [1963] and Murphy [1972, 1973] exposed an interpretable separation of BS_N into the sum

$$N^{-1} \sum_{j=1}^{k} n_j(P_j - (r_j/n_j))^2 + N^{-1} \sum_{j=1}^{k} n_j(r_j/n_j)(1 - (r_j/n_j)),$$

denoted as $C_N + R_N$. Here k is the number of distinct values of probability assessments in the sequence, N is the total number of indicators, n_j is the number of those with assessed probability P_j, and r_j is the corresponding number of those n_j indicators which are observed to equal one. The first term, C_N, has been interpreted as a measure of the probability assessor's calibration, scoring the extent to which the observed proportions r_j/n_j differ from the corresponding assessed probabilities P_j. The second, R_N, is considered a measure of the assessor's refinement, scoring the extent to which the distinct previsions distinguish homogeneous or heterogeneous groups of indicators. (See DeGroot and Fienberg [1983].)

Now consider the single random quantity X with the assessed n-tiles $X_{(1)}, \dots X_{(n)}$ as before, and cumulate the Brier score over the $N = \sum_{j=1}^{n} {}_nC_j = (2^n - 1)$ different indicators with assessed probabilities implied by the n-tile specification.[1] Thus $k = n$, and $n_j = {}_nC_j$. The theorem proved above implies that the calibration assessment term, $N^{-1} \sum_{j=1}^{k} n_j(P_j - (r_j/n_j))^2$, equals zero for every coherent assessor. For when all those specified assessments of events are scored, $P_j = r_j/n_j = j/n$ for every $j = 1, \dots, k$. Moreover the refinement term, $N^{-1} \sum_{j=1}^{k} n_j(r_j/n_j)(1 - (r_j/n_j))$, is also identical for every assessor, equal to $(2^n - 1)^{-1} \sum_{j=1}^{k} {}_nC_j\,(j/n)\,(1 - (j/n))$. Thus this global Brier score is identical for every coherent n-tile specification.

While this score is identical for every assessor, it must be remembered that the specified n-tiles (on the basis of which the various indicators are defined) are different for each assessor. Although the 'global' Brier score for the assessment of X analysed above is identical for every assessor, the meaning of refinement as generally understood is washed out since one person's n-tiles could be quite closely bunched while another's could be well spread out. The surprise of apparently identical refinement scores for two arbitrary distributions disappears when it is recognised that the expected quadratic loss, $E[X - E(X)]^2$, is clearly different for such different persons.

This analysis is provocative in that it specifies exactly the sense in which any coherent assessed probability distribution is 'just as good' as any other.

A final note should be made concerning so-called 'good calibration' and the Brier score when some proper subset of all elicited previsions is scored. Over any such subsequence, both the best (lowest) as well as the worst (highest) Brier score will be achieved by assessors who are uncalibrated according to the currently honoured frequency criterion. Suppose a sequence of previsions for independent indicators is scored, the prototypic case for the calibrationist arguments. The best possible Brier score will be achieved if $I_i = 1$ whenever $P_i > \frac{1}{2}$ and $I_i = 0$ whenever $P_i < \frac{1}{2}$. Similarly the worst score would be achieved if $I_i = 0$ whenever $P_i > \frac{1}{2}$, and $I_i = 1$ whenever $P_i < \frac{1}{2}$. Each of these conditions implies that the refinement score $R_N = 0$, and that thus $BS_N = C_N$. No matter how large N may be, the best (and the worst) Brier scores will be achieved by uncalibrated (in the relative frequency sense) assessors! A graphical exposition of the local Brier score, calibration, and refinement which illuminates this result is presented in Blattenberger and Lad [1984].

---

[1] We might call this a 'global' Brier score of all the previsions implied by coherence from the specification of those n-tiles. Similarly, the Brier score of any subset of these $(2^n - 1)$ previsions might be called a 'local' Brier score.

5 ON INCOHERENCE AND LUCK

Why do some people flourish and others not? Is it solely because some are well informed and others not, or can luck be properly considered to influence the progress of one's own state of affairs? For some time I had rejected the notion that a consistent subjectivist position could refer to luck in explaining the outcome of day to day affairs. For this seems to hinge on assuming that the state of the world is the outcome of chance phenomena. That well formed notions of luck are reducible to chance was a conclusion argued by De Moivre [1756, pp. iii–v].

The analysis of the previous section, however, pushes to the fore a sense in which a subjectivist might well consider him/herself lucky or unlucky in evaluating his/her own knowledge and experience. In the course of life, no person actually engages in all the bets which he/she might be willing to take. The sequence of actions actually taken by any particular person is one of a whole collection of sequences which he/she would be equally disposed to take. If the sequence taken results in unfavorable experience for the actor, one can defend as coherent the state of knowledge upon which he/she acted by pointing to other sequences of acts (bets) toward which he/she had been equally disposed and which would have yielded preferable outcomes. Concomitantly, in the face of unfortunate experience, one ought to search one's opinions for encumbered but unelicited incoherency. Opinions which would make one a sure loser, if all were expressed, would probably make one a loser if partially expressed. This conclusion is a consequence of the fact that incoherent assessments are not well-calibrated according to the definition proposed in this essay.

This two dimensional response to unfortunate experience is the proper reaction supported by the theory of coherent probability. It contrasts directly with the response demanded by Fréchet and too many calibrationists—a castigation of the validity of one's originally specified probability assessments.

The pursuit of calibration analysis yields no substantiation for the concept of a 'correct' probability assessment. When calibration is considered as a global property of the entire belief distribution, then every coherent specified assessment has been shown to be well-calibrated.

University Of Utah, Salt Lake City

REFERENCES

BLATTENBERGER, G. and LAD, F. [1984]: 'Separating the Brier Score into Calibration and Refinement Components: A Graphical Exposition', Department of Economics Research Report, University of Utah.

BRIER, G. W. [1950]: 'Verification of Forecasts Expressed in Terms of Probability', *Monthly Weather Review*, 75.

DAWID, A. P. [1982]: 'The Well-Calibrated Bayesian', *Journal of the American Statistical Association*, 76.

DE FINETTI, B. [1974]: *Theory of Probability*, vol. I. John Wiley and Sons.

DE FINETTI, B. [1962]: 'Does it Make Sense to Speak of "Good Probability Appraisers"?', in *The Scientist Speculates: An Anthology of Partly Baked Ideas*, edited by I. J. Good. Heinemann.

DEGROOT, M. and FIENBERG, S. [1983]: 'The Comparison and Evaluation of Forecasters', *The Statistician*, 32.

DE MOIVRE, A. [1756]: *The Doctrine of Chances*, Third Edition. London. Facsimile reprint of 1967. Chelsea.

FRÉCHET, M. [1955a]: *Les Mathématiques et le Concret*. Presses Universitaires de France.

FRÉCHET, M. [1955b]: 'Sur L'Importance en Econometrie de la Distinction entre les Probabilités Rationelles et Irrationnelles', *Econometrica*, 23.

FRÉCHET, M. [1949]: 'Au Congrès International de Philosophie Des Sciences: Rapport Général sur les Travaux de la Section des Probabilités', reprinted in *Les Mathématiques et le Concret*.

FRÉCHET, M. [1946]: 'Les Definitions Courantes de la Probabilité', *Revue Philosophique*, reprinted in *Les Mathématiques et le Concret*.

KADANE, J. and LICHTENSTEIN, S. [1982]: 'A Subjectivist View of Calibration', Carnegie Mellon Department of Statistics Technical Report #233.

LICHTENSTEIN, S., FISCHOFF, B. and PHILLIPS, L. [1982]: 'Calibration of Probabilities: The State of the Art to 1980', in D. Kahneman, P. Slovic and A. Tversky (*eds.*), *Judgment under Uncertainty: Heuristics and Biases*. Cambridge University Press.

LUCAS, R. [1973]: 'Some International Evidence Concerning Output-Inflation Tradeoffs', *American Economic Review*, 63.

LUCAS, R. and SARGENT, T., editors [1981]: *Rational Expectations and Econometric Practice*. University of Minnesota Press.

MURPHY, A. H. [1972]: 'Scalar and Vector Partitions of the Probability Score (Part I). Two-State Situation', *Journal of Applied Meteorology*, 11.

MURPHY, A. H. [1973]: 'A New Vector Partition of the Probability Score', *Journal of Applied Meteorology*, 12.

PRATT, J. [1962]: 'Must Subjective Probabilities be Realized as Relative Frequencies?', unpublished seminar paper, Harvard University Graduate School of Business Administration.

SANDERS, F. [1963]: 'On Subjective Probability Forecasting', *Journal of Applied Meteorology*, 2.

SAVAGE, L. J. [1954]: *The Foundations of Statistics*. Second revised edition 1972. Dover Publications.

Brit. J. Phil. Sci. 35 (1984), 223–243 *Printed in Great Britain*

Is Consciousness Important?

by KATHLEEN V. WILKES

The paper discusses the utility of the notion of consciousness for the behavioural and brain sciences. It describes four distinctively different senses of 'conscious', and argues that to cope with the heterogeneous phenomena loosely indicated thereby, these sciences not only do not but should not discuss them in terms of 'consciousness'. It is thus suggested that 'the problem' allegedly posed to scientists by consciousness is unreal; one need neither adopt a realist stance with respect to it, nor include the term and its cognates in the sciences' conceptual apparatus. The paper briefly examines Nagel's [1974] article, since this presents the strongest counter to the thesis proposed.

1 I begin with four quotations:

(*a*) '"Consciousness" is the name of a nonentity, and has no right to a place amongst first principles' (James [1912], p. 2);

(*b*) 'To anyone but a psychologist, it would seem incredible that textbooks on thinking can ignore or say little about the stream of consciousness' (Pope and Singer [1978], p. 102);

(*c*) 'What is meant by consciousness we need not discuss; it is beyond all doubt' (Freud [1964], p. 70);

(*d*) 'Consciousness is like the Trinity; if it is explained so that you understand it, it hasn't been explained correctly' (Joynt [1981], p. 108).

These well illustrate the problem I wish to discuss. James, who affirmed that 'consciousness' names a nonentity, notoriously went on to write eloquently about the stream of consciousness, as Pope and Singer require; Freud and Joynt seem too far apart to be able to argue. What should the brain and behavioural sciences do about the notion of consciousness?

2 The decline of radical behaviourism has freed psychologists to take the mental seriously once again. Simultaneously, it has made psychology appear to have concerns close to those that interest the philosopher of mind; both disciplines, after all, discuss such phenomena as memory, emotion, belief, desire, pain, sensation and so forth, and one might hope that each would benefit from this renewed coexistence.

The apparent legitimation of the study of run-of-the-mill mental phenomena such as those mentioned above is not enough, however, for many psychologists and neuroscientists. As Everest was to mountaineers

Received 11 February 1983

before 1953, so, for some, consciousness is the yet-to-be-mastered problem; until that nut is cracked, psychology and the brain sciences remain seriously incomplete. The concern with consciousness is now to be seen wherever you look, but typically in significant places: in the *titles* of collections, in Presidential Addresses summarising in general terms the state of the discipline, or, when a textbook is consulted, usually either in the Preface or in the concluding chapter. The notion crops up rather infrequently in the bread-and-butter reports of work in experimental psychology. Philosophers interested in the work of their laboratory colleagues share this ambivalent attitude to the problem; some, like Nagel (Nagel [1974, 1979]), suggest that it is in principle impossible for an objective science to do justice to subjective experience, while others, like Dennett (Dennett [1978]), prod cognitive psychologists along by sketching an outline theory of consciousness within which research into the phenomenon might proceed.

In this paper I shall question the existence of a 'problem of consciousness', suggesting that in fact consciousness as such is not at all important, and that psychology and the neurosciences would lose nothing, and gain much, by refusing to chase this will-o'-the-wisp. It should, though, become clear that this is not a particularly radical suggestion. It is much tamer than a more extreme argument to the effect that commonsense psychology and the philosophy of mind need not bother with the notion either; I shall conclude by defending briefly this further claim, but the bulk of the paper is concerned with the scientists' problem.

3 Nobody, I assume, would wish to argue that 'conscious', or its nominalisation 'consciousness', are clear notions. The range of disagreement about the scope of the terms puts that out of the question—on the one hand we find those liberally expecting to ascribe consciousness one day to computers, taking it for granted that human and many infrahuman animals are indubitably conscious; while on the other we find those like Sir John Eccles, or Descartes, who do or might deny consciousness to artefacts, to the right hemisphere of the human brain, and to non-human animals.

Moreover, 'conscious' as an adjective is ascribed to a pretty heterogeneous set of things, and it has at least four distinguishable uses. In one of them 'conscious' is a one-place predicate which takes complete systems (organisms, and maybe robots) as subject: John is conscious now, Jessie is not yet conscious. The remaining three uses all treat 'conscious' either as a two-place, relational term: John is conscious of Jessie's grimace; or as a one-place predicate taking mental states or events as subjects: conscious beliefs, conscious perceptions. The forthcoming list is not, however, intended as an exclusive or exhaustive botanization of types of consciousness (indeed, if ways of being conscious were so easily classifiable, the notion would be more tractable than I believe it to be). The intention is rather to underline the heterogeneity of conscious phenomena, to suggest that there are *at least* four interestingly distinguishable kinds of ascriptions of consciousness.

The first of my four is the use of 'conscious' as a one-place predicate of individuals. Here 'conscious' means, roughly, 'awake' as opposed to being asleep or in a coma. It is common to discount this use of the term, regarding it as elliptical for expressions taking the form 'conscious of' or 'conscious that'; but it might seem that in its own right it should interest psychology and the neurosciences, inasmuch as these disciplines certainly need to explain the nature of, and differences between, sleeping and waking states as such—and also, of course, they have to take on as *explananda* states of dreaming, hypnosis, fugue, epileptic automatism, and the like. In any case, whether we have here a primary or a derivative sense of the term, this use of 'conscious' looks easy to understand and need not detain us yet.

The remaining three all admit both a relational use ('*A* is conscious of *x*') and a non-relational use in which 'conscious' is a predicate of mental states and events (pains, beliefs, perception). Of the three, perhaps the most familiar is the use whereby the mental phenomena to which 'conscious' might be ascribed comprise the set of sensations. Now sensations have traditionally been picked out in the philosophical literature by various epistemological features that are said to characterise our relationship to them: they are private to their possessors, and we know of them immediately, without inference—a fact which gives us not only an asymmetrical but indeed a privileged access to them, so that our judgements about them are incorrigible. We could summarize all this by saying that the *esse* of sensations is *percipi*. However, it would be as well to supply non-epistemological characteristics of sensations, since all these features have been disputed. The ground is familiar: 'privacy', for example, has a number of distinct senses itself (Ayer [1963]); one interpretation collapses 'private' into 'what we are incorrigible about', another gives it the unconvincing reading 'incommunicable', while another sense is the innocuous 'inalienable', whereby you can no more have my pain than you can have my birth-certificate. Knowledge by acquaintance has been attacked by a flourishing tradition which denies that there is awareness or knowledge of anything except under a description (see Shoemaker [1975], who does not however cast the argument in these terms); both 'immediacy' and 'non-inferentiality' become dubious when we allow (as we surely should, and as psychologists must) for the existence of tacit knowledge and subdoxastic states; incorrigibility is either false, or needs to be weakened (*à la* Rorty [1970*a*]) to something like: 'at present, the testimony of the subject is taken to be authoritative'. (Indeed, it would not be unfair to abandon as inadequate or unclear the epistemological marks of mental phenomena; they were, after all, introduced *not* to provide a perspicuous characterisation of the mental, but rather to pick out a class of statements immune to the Cartesian sceptic. These days we tend no longer to attempt to rebut scepticism by appeal to a privileged set of indubitable propositions, thinking that we have other means of handling the problem; hence the alleged epistemological features of some mental phenomena, sensations in particular, may be replaced too. (For a fuller defence of this line of argument, see Rorty [1970*b*]).) Fortunately, it is not difficult to delineate the set of sensations by examining paradigm instances: pains, itches, tingles, butterflies in the stomach, pins and needles, tickles, twitches. From consideration of these we can see that sensations are occurrent and clockable—they have duration, and we can often count them—that they are located, often quite precisely, and that they all fit somewhere on one or more of a range of qualitative spectra: from faint to intense, from pleasant to unpleasant, from steady to spasmodic, and so forth. Evidently, we need not ignore the element of truth that is roughly captured by the epistemological features listed above, since we can acknowledge that their *esse* is indeed *percipi*—simply because nothing is allowed to count as a sensation unless it is perceived to be such: that is the way our linguistic habits have it (and that is why sensations are the *only* mental phenomena that firmly fit the epistemological criteria, for it is only with these that our linguistic conventions are thus strong). It might, or might not, be convenient for a scientific psychology to adopt the convention that the *esse* of sensations is *percipi*; that is as yet an open question.

The third use of 'conscious' (like the second, either used relationally, as in '*A* is conscious of *x*', or as a predicate of mental phenomena) is typically, and I believe erroneously, fused or confused with the second. I mean, of course, its use to characterise our sensory experience (which should include kinaesthesia along with the five senses). The confusion is expressed in, and illustrated by, the fact that philosophical terms of art such as 'quale', 'sense datum' or 'raw feel' are employed to pick out both sensations of pain and the experience of seeing something red. That such a fusion is mistaken is easy to argue. Sensory experience is not jointed—it does not occur in the manner of beads on a string. It is not locatable (we see *with* our eyes, but this is not where the seeing *is*); it has no duration, except in the boring sense that we can stop seeing by closing our eyes and start again by opening them—it is less easy to suppress the other senses. We cannot sensibly count visual or auditory experience*s*; in other words, it is difficult to make much sense of the plural of 'experience' where the five senses are concerned. Sensory experience is usually not set on qualitative spectra (such as faint-intense, pleasant-unpleasant, steady-spasmodic) except in a parasitical way—the partially blind or partially deaf may experience clear objects and loud noises as 'faint' or 'indistinct', and to look at a Monet may be a pleasant visual experience; but except in such cases as these, qualitative properties fail to apply. Moreover, if we resurrect for a moment the epistemological features of some mental phenomena, here too sensory experience is clearly distinguishable from the having of sensations; it is, presumably, now impossible to overlook the mass of evidence for subliminal perception, and 'blindsight', whereby a subject with a specific lesion is able to discriminate shapes, colours and patterns while denying that he sees anything (see Weiskrantz *et al.* [1974], and Weiskrantz [1980]), has provided another kind of example. The *esse* of sensory experience is emphatically not *percipi*, nor are we in any important sense privileged about, or incorrigible with respect to, it. (The alleged incorrigibility of 'I have a Ø-experience' is of course no counterexample; the content of this is exhausted by 'I seem to see something Ø', and this is an incorrigibility with respect to thought rather than with respect to sensory experience; we could, after all, play the same game with 'a walking experience'/'seeming to walk'. I shall be discussing thoughts shortly.)

It is difficult to understand exactly what we are doing when we call some occasions of sensory experience 'conscious' and others not. Presumably we can say that subliminal perception and blindsight count firmly as non-conscious; what is far less clear is whether run-of-the-mill sensory experience, when we successfully get around even if preoccupied with something other than the immediate environment, should be called conscious. The tempting test is 'ability to report': we are conscious of the sensory experience that we can describe. But such a criterion is in many respects suspect. It favours that subclass of conscious systems which have the ability to use language, leaving us in the dark about how to sort out conscious from non-conscious experience in infrahuman animals; more importantly, perhaps, it delivers highly ambiguous results even when language-users are in question. For, if we ask an agent to report upon his contemporaneous visual or auditory experience, then we have altered the phenomenon: someone concerned to describe his current sensory input is behaving very differently with respect to it than is the man who, preoccupied with his forthcoming dinner party, negotiates obstacles in a supermarket. But if we rely instead upon *ex post facto* reports, then the results are inconclusive. The hackneyed example of a driver keeping up a flow of conversation while driving skilfully is all we need; he may be unable, even after a tiny time-lag, to say anything at all about the nature of the road over which he has just driven. Yet evidently his sensory discrimination has been subtle and refined, certainly not subliminal; in fact we would expect to find subliminal and non-subliminal distinctions in the perceptual activity that helps to explain his driving skill. Much more of our routine sensory experience is like this than we perhaps care to realise; it escapes our attention, maybe, because when we consider and concentrate upon our sensory experience, we alter the phenomenon we sought to describe.

At best, it seems, we must make do with clear cases of conscious sensory experience—as, for example, when a subject is reporting upon the current contents of his visual field—and clear cases of non-conscious experience—as, for example, in blindsight or in subliminal perception. To resort to this way of marking the distinction is, however, to admit that the bulk of our sensory experience (including the most common and interesting forms), and practically all the sensory experience of other animals, elude or resist the dichotomy.

The last category of mental phenomena to get prefixed (or not prefixed) by the adjective 'conscious' is itself heterogeneous; it includes all the propositional attitudes. Two distinctions need to be made in this large class.

The first is that between occurrent and dispositional attitudes. Thoughts can suddenly strike one, tunes may run through the head, even to the point of distraction; one may engage in explicit deliberation, thinking in words, and memories may unexpectedly flood back (consider Proust's experience with the madeleine cake). On the other hand there are the beliefs, desires, memories, aversions, preferences (*etc.*) which we can ascribe fairly to the sleeping man, or to an animal. The distinction between occurrent and dispositional propositional attitudes does not, however, tally with a conscious/non-conscious division. For whereas the most familiar instances of occurrent thoughts are typically regarded as conscious, there is no incoherence to the idea that pre-, sub-, or unconscious occurrent thoughts may exist. Indeed, even the treacherous test of reportability cannot be deployed here, since thoughts which we should surely consider to be preconscious can be directly reported: one who stumbles at the top of a flight of stairs may say, as he dusts himself down, 'I thought there was another step'—and his testimony is surely privileged to the same extent (whatever that may be) as is that about thoughts explicitly running through the head. Turning to the dispositional half of the dichotomy, we seem to talk sensibly about the conscious beliefs of a sleeping man, or of someone engaged in an activity entirely irrelevant to the belief ascribed; nor are such ascriptions elliptical for statements to the effect that the agent often, or indeed ever, thinks explicitly of the proposition which we say he believes—it may, for instance, be a proposition that follows easily from other propositions we know he holds true, but one which he has never made explicit to himself. Perhaps an ascription of a conscious belief is elliptical for something like 'if you were to ask him whether he thinks that *p*, he would assent', but here again we have altered the phenomenon we sought to describe by suggesting that the subject's attention be drawn to the thought in question; and, moreover, linguistic behaviour is only one of the many kinds of behaviour that we use to justify our ascriptions of belief, so that it seems somewhat arbitrary to favour it (as well as being useless when we need to consider animals).

The second subdivision of the propositional attitudes is the distinction between those that take, and those that do not take, the agent himself or his own mental states as their subject-matter. This distinction evidently correlates reasonably well with that between self-conscious thoughts (*etc.*) and the rest (and see Frankfurt [1971]). Interestingly enough, it seems that self-conscious attitudes need not themselves be conscious. That is, someone can have beliefs about himself that, intuitively, seem not conscious. One example might be the anorexic who appears to have a non-conscious belief about her body being fatter than it evidently is; and in an ingenious experiment in which a number of subjects heard a tape of twenty different voices reading the same innocuous sentence, their own voice being the twelfth recorded, various physiological tests (such as GSRs, EMGs (electromylograms), and finger plethysmograms) showed in almost all cases clear autonomic arousal reactions to their own voices, even though more than half failed to recognise any voice, let alone the twelfth, as their own (Holzman, Rousey and Snyder [1966]). No doubt self-consciousness is in its way as vexed and obscure a notion as is that of consciousness; however it seems not unambiguously to be a clear, if sophisticated, kind of consciousness.

To return to consciousness and the propositional attitudes in general: while, again, we have clear paradigm cases of admittedly conscious thoughts—*e.g.*, when a sentence runs through the head—and clear paradigm cases of non-conscious ones—*e.g.*, the disavowed beliefs of the self-deceived, or the beliefs of the Freudian Unconscious—the vast majority once more evade the dichotomy.

4 From the preceding section, consciousness emerges not only as thoroughly heterogeneous, but also as a *prima facie* unpromising phenomenon for systematic exploration. The majority of psychologically interesting or important events, states and processes seem not clearly conscious, but not evidently non-conscious. I shall shortly go on to discuss what the brain and behavioural sciences are doing, and can be expected to do, with respect to sleeping and waking, sensations, sensory experience, and the propositional attitudes; but first I want to mention, in order to reject, two attractive intuitions about the nature of consciousness which, if not expelled, might confuse the argument.

The first such intuition is that consciousness is a kind of internal illumination; a spotlight in the private inner theatre of the mind that picks out some, but not all, of the passing show; a light that may be dimmed or intense, loosely or narrowly focussed, but which is unambiguously either on or off. This metaphor, though, will not suffice; it conflicts with many common uses of the term 'conscious'. Consider sensory experience. We can and do say that we are conscious of things, but more often than not these are public phenomena—the highlights in a painting, the expression on a face, the rumble of traffic outside. The technique of attending to one's experience as such is a sophisticated and rarely-used one. Nor do sensations fit the 'spotlight' model; we seem to regard it as one and the same thing to have a sensation and to be conscious of that sensation—illumination and illumined coalesce. Similarly, conscious thoughts are not: thoughts about the external world *plus* an inner apprehension of the having of such thoughts; were that to be the case, rather few normal thoughts would be conscious. In fact, the inner spotlight idea, if it grips at all, looks more appropriate to conscious *self*-consciousness, where we can make sense of the idea that a conscious thought takes as its object a mental event of some kind; the metaphor here is not wholly misleading, and presumably lies behind the idea of *introspection*. But apart from that, the spotlight analogy, deriving as it does from the 'private inner theatre' metaphor, must be rejected along with that powerful, but powerfully misleading, picture. However we are to understand the nature of consciousness—if indeed it is the sort of thing that *has* a nature—it will not be in terms of this on-off, yes-no model of an internal illumination.

The second intuition to guard against is closely related to the first, in that spotlights can be focussed or unfocussed; it is the idea that consciousness is some form of attention. Attention, whatever else may be said of it, is among other things the allocation of our cognitive resources, and such allocating is as critical when subdoxastic states are in question as it is when we are attending consciously. Moreover, 'any problem-solving or game-playing computer pays attention . . . first to one candidate course of action and then to another, and presumably it would not on this ground be deemed conscious' (Dennett [1978], p. 209). In fact, even if consciousness were helpfully to be understood in terms of attention, we would not get much forrader, because the same problem would plague the ascription of attention in the examples discussed above: to what, if anything, is the man in the supermarket, who is preoccupied with his forthcoming dinner party, attending? To what extent must we uninterruptedly attend to our pains if we are truly to have them? Some attention is conscious, and some consciousness is attentive; more than that we cannot usefully say. The main point to note, though, is that since attention can be clearly non-conscious, we cannot hope to understand consciousness by assimilating consciousness to attention.

5 I return now to the rough list of four types of conscious phenomena, to consider what psychology and the neurosciences are doing and can do about them.

The first was the distinction between consciousness as a waking, and non-consciousness as a sleeping or comatose, state. It goes without saying that the distinction between sleeping and waking constitutes one *explanandum* for science, and it is a lively field of study; nevertheless, this is only one element of a more exhaustive investigation of these and a variety of other states. Sleeping itself divides into REM sleep and several different levels of NREM sleep, and physiological considerations suggest that some at least of the various sleeping states should more profitably be likened to the waking state than to other sleeping states. REM sleep is tied to the dreaming process (although some rather fragmentary dreams seem to occur occasionally during NREM periods), and dreams evidently present a major topic for systematic examination—yet the conscious/non-conscious dichotomy breaks down completely here. To some, dreaming seems a way of being conscious, to others it is patently a form of non-consciousness. There are clear reasons for calling dreams conscious phenomena; after all, one has thoughts, perceptions, plans and surprises in dreams (or so it seems). On the other hand the dreamer usually lies placidly, immune to the environment—the very epitome of unconsciousness. I shall later suggest that there is no fact of the matter here at all, no true answer to whether the dreamer is conscious or not; but meanwhile we should note that dreaming is not the only difficult sleeping state for the conscious/non-conscious dichotomy to grip: sleep-walking and sleep-talking also present problems for it. Both typically occur primarily during NREM sleep, but each indicates mental processing of a highly sophisticated kind: the sleep-talker talks, and linguistic behaviour is usually taken to be the most reliable sign of consciousness; while the sleep-walker negotiates obstacles and engages in often reasonable and directed activity. More complex behaviour yet can be seen with those in a state of epileptic automatism. One sufferer, a doctor, while in such a state conducted a reasonably efficient examination of a patient, measuring blood-pressure, heart rate, pupil dilation and so forth, and was even able to write up the examination in his casebook; the question whether he was conscious or not surely just does not arise here—*that* is not what is interesting about the case (and fortunately not, because we have no idea what to say in those terms). Presumably fugue states qualify as conscious, if only because they may last months or even years, and deceive any onlooker; we would take the Rev. Ansel Bourne, who in his fugue period spent three months as a small shopkeeper, to have been conscious during that time—he had to weigh the groceries, count the change, converse with his customers, *etc.*—despite his total, if temporary, amnesia for his clerical life and despite the fact that when restored to normality he was amnesic for his shopkeeping period (see James [1890], pp. 391–3). However, hypnotic states may be far harder to classify. There are several depths of hypnotic trance, all of which have some features in common with waking, or with REM sleep, or with one of the several forms of NREM sleep; both behavioural and physiological evidence testify to this. In sum, classification in terms of consciousness or its absence is simply too crude to cope with all this diversity; surely profitable research into these and other phenomena will require a theoretical classification determined along different, and probably more fine-grained, principles.

Turning next to the second form of consciousness, the experiencing of sensations, much the same holds true here too. I shall run the argument in terms of the sensation of pain (as do most when discussing sensations; pain is certainly one of the most important and striking of the class, and it is doubtful that the discussion would differ significantly if thirst, say, or tickling were examined instead). It is indeed important to insist that in humans at least most paradigm instances of pain are inseparable from the knowledge or belief that the subject is in pain. This feature, which cast in different terms just is the consciousness of pain, is thus something which any adequate account of pain must accept as an *explanandum*.

There seems, though, no reason of principle why psychology and the neurosciences should be debarred from taking this type of consciousness into account, even though it may not be *qua* consciousness that they do so. Consider: the line of research that looks at present to be the most profitable treats pain as a phenomenon with a causal-functional role in the psycho-physical economy of the organism. The occupant of this causal role cannot, despite the handwaving abbreviations of some philosophers, be sensibly regarded as a single state (*e.g.*, 'C-fibre firing'), but should rather be treated as a set of states, running both sequentially and in parallel, of which set the stimulation of C-fibres in the brainstem is usually an early member. The states themselves may be described in functional or—as our knowledge permits—in structural terms as well; this does not matter for our purposes.[1] The sequences of states initiated by C-fibre activity are considered by some contemporary theoreticians to be series of information-processing states, and the processing of such information may require the postulation of (*inter alia*) hierarchically-ordered sets of processors. Thus it may implicate not only the paleocortical aversive-motivational system, but also a hierarchy of stages in the neocortical, more cognitive system (*cf.* Melzack and Wall [1970]). These complex sequences of processes may have familiar behavioural outcomes: the sufferer typically resiles from the source of the pain, tends the injured area, screams, telephones a doctor, hunts for iodine and bandages. Very frequently the complexity of the behavioural response appropriate to the pain for that sufferer will presuppose a corresponding complexity in the psycho-physical processes postulated by the theory; for instance, for an adult human to take proper care of a burn or cut requires that he has beliefs and desires about the pain, its extent, and its intensity, *i.e.*, it requires the activity of the higher cognitive functions. That this is not invariably required can be seen if we consider that there are cases in which we find that we have snatched a hand from the source of pain *before* we knew what was going on—before, that is, the relevant information was assimilated by the 'higher' processors—even though the explicit recognition of pain will speedily ensue.

This last observation leads in to a much neglected, but highly significant, form of (mild) pain for which the theoretical explanation remains basically the same although the information may not reach the sophisticated cognitive levels; I have in mind the mild and unnoticed discomfort of a slightly overstrained muscle, for example, or of low degrees of strain, pinching or pressure. Throughout the day, and indeed throughout the night too, we make hundreds of minute postural adjustments to alleviate or prevent such discomfort, and we are usually not aware of doing so—when we are asleep we are emphatically not aware of doing so. That such states belong on the same spectrum with paradigm instances of pain is something readily discovered by anyone who is for some reason prevented from making the tiny shifts of position required for their elimination. The congenitally pain-insensitive are not merely individuals who do not *feel* pain consciously, but are rather people who simply do not *have* pain, not even these mild and unnoticed forms—there is nothing with the relevant causal-functional role going on in them. This is borne out by the fact that it is not only cuts and burns to which they fail to respond, but also strains and pressures to which normal people adjust non-consciously; for they are not only in danger from untended burns, breaks and cuts but also tend to end up warped and lamed because they cannot compensate for muscular strain (the proper care of such individuals requires that they be regularly moved about when they are in bed, and encouraged to shift position frequently when awake). Thus there is excellent reason to talk of unnoticed pain; because the response required to deal with minor strain or pressure is so easy—uncrossing a leg, shifting slightly the orientation of the head—there is no reason why the strain or pressure should be consciously recognized as such: we can, that is, manage without consciousness, higher control, here.

---

[1] Moreover, it is often arbitrary and uninteresting whether something is to be classified as a 'functional' or as a 'structural' description; this may reflect only the role for which that description is being used at a specific level of theory—consider how the terms 'comparator', or 'stimulus analyser' may fluctuate in import from 'functional' to 'structural'. The distinction is *far* less fundamental than many philosophical functionalists suppose.

One of the pressing problems for the brain and behavioural sciences is the construction of a taxonomy of *explananda.* It is evident that pain, under some description, must feature in such a taxonomy. However, what counts as pain for the purposes of the sciences may prove to be in part at least a product of what kinds of phenomena the *explanans* serves to unify, and thus may diverge from the extension of the class of phenomena picked out by the vernacular term. I am suggesting that examination of the functional description of pain may, and probably will, require that these instances of unnoticed pain count, properly and equally, as pain discrimination by the organism despite the fact that we need not be aware either of the painful stimulus or of the compensatory response; and hence that the sciences may for principled reasons refuse to agree with the vernacular that the *esse* of pain is *percipi*. If so, there would be nothing incoherent in the study of pain discrimination in infrahuman animals by someone who is as agnostic as is Eccles about the question of their consciousness; pain would no longer *presuppose* the transparency of immediate awareness. Nonetheless, pain-research so conceived need omit nothing, for it remains a fact that the most interesting and important cases of pain in humans (and *I* would say in the higher non-human animals as well) indeed implicate the higher cognitive processes, without which appropriate responses to noxious stimuli would often be difficult or impossible. These processes then need to be described and explained; though there are colossal difficulties in practice, there seems no bar in principle to such a psycho-physical endeavour. It would be perfectly legitimate to describe such a project as research into the consciousness of pain; legitimate, but not particularly helpful, since the most perspicuous level of description would count it as research into the highest of a series of stages of processing of a certain range of stimuli.

Furthermore, discussion in terms of consciousness often obfuscates and confuses one's ability to describe and explain other sorts of pain. Take, for example, hypnotic anaesthesia. A typical experiment is the following: subjects are hypnotized, told they will feel no pain, and then one arm is put into a stream of circulating iced water. This is rapidly experienced as unpleasantly painful by the unhypnotized; but subjects under hypnosis may sincerely report that they feel no pain, and will leave their arm in the water for long periods, apparently untroubled. On the other hand, and *with* the other hand—if it is supplied with pencil and paper—the subject typically provides a simultaneous running complaint about the intensity and unpleasantness of the pain (see Hilgard [1977], pp. 135–203). Such forms of dissociation, common enough in hypnosis and elsewhere, make nonsense of the attempt to describe what is going on in terms of consciousness or its absence. Nor seems there to be a fact of the matter about whether the hypnotized subject is *really* in pain or not. After all, the vernacular concept of pain is a pretty crude notion which includes a heterogeneous diversity of sensations (few sensations are less similar than a dull stomach ache and the prick of a pin); it is only to be expected that a scientific taxonomy should delineate the extension rather less crudely and—perhaps—in accordance with theoretical demands of its own which may include unnoticed and dissociated pain as legitimate kinds.

Mutatis mutandis, the argument adjusts readily to fit our third category, that of sensory experience. Indeed, it will fit more easily, inasmuch as we seem far less firm in our intuition that sensory experience must always or even primarily mean *conscious* experience. I suggest again that consciousness as such will not figure as a unique *explanandum*, even though what we are trying to do with the prefix 'conscious' should get adequately acknowledged in the scientific account. Once more the information-processing model dominates much of the research in this domain (or, what amounts to the same thing for our purposes, theories citing degrees and grades of responsiveness); and there will be, or we should expect to find, important and systematic differences between perception classified as subliminal, and perception of what occupies the central focus of our visual field. Such a distinction, as we have already noted, fails to tally well with any non-conscious/conscious dichotomy, but they overlap up to a point at least; to the extent that they do overlap, the differences in the accounts will represent an explanation of the differences between conscious and non-conscious perception.

The vexed problem of qualia—if one takes seriously the existence of sensory qualia or 'raw feels', as I admit I find it hard to do—is entirely irrelevant to the argument. That it is irrelevant is surely shown by the fact that qualia, if they are to be postulated at all, must equally be postulated at a non-conscious level; the role they are ascribed is that of explaining how we distinguish objects on the basis of their colour (or other sensory) qualities, but subliminal perception too involves colour discrimination, and D.B. (Weiskrantz's first blindsighted human subject) was able to report the colours of stimuli flashed to the blindsighted part of his visual field—although, of course, D.B. thought he was guessing randomly. Hence, *if* qualia are postulated to explain colour discrimination, we would need them at the non-conscious levels too. One could of course stipulate that qualia can be ascribed only when conscious (or at least not subliminal or blindsighted) perception is the subject of discussion—after all, 'quale' is a term of art that can be legislated to mean whatever one wants. Even so, no particular or novel problems arise. For we shall still need to investigate the psychophysical processes implicated in ordinary ('conscious') perceiving, in subliminal perception, and in blindsight; and whatever characterizes the first but not the last two cases we can, *if we wish*, call the functional basis of qualia.

Dropping qualia: the postulation and study of more and less complex and hierarchically-ordered levels of information-processing should be expected to explain (*inter alia*) our ability to recognize explicitly (approximately: to be conscious of) objects of perception. However difficult—and the understanding of higher-order cognitive processing is at present extremely sketchy—the problem is not, or seems not to be, one of principle.

A reasonably common objection to such an account is that it would allow that computers are, or could soon become, genuine subjects for the ascription of sensory experience; yet computers are surely insentient and non-conscious. Such an objection is, I believe, confused. Intuition is a frail prop here. We have no clear intuition about whether the frog, which flicks out its tongue at small moving black objects crossing its visual field, is 'conscious' of the flies it catches; if we are thus uncertain about frogs, it is difficult to defend the ban on all forms of machine sentience. Moreover, not all by any means share this anti-mechanistic intuition. Second: the causal/functional role which paradigm instances of perception display, and which determines what it is to be a state of perceiving, is vastly richer than that open to any computer, existing or contemplated. When humans see things, they may paint them or write poems about them, ignore, avoid, move or eat them, sit on them or bomb them; infrahuman animals have a somewhat more restricted, but still extensive, range of ways of interacting with perceived objects; the computer, though, like the frog, is severely limited in what it can do with 'seen' objects—hence, of course, the legitimation of the scare-quotes around 'seen'. Computer 'perception' shares too little of the causal/functional role that helps define animal perception to justify removal of the scare-quotes; and, although the difference may be one of degree rather than kind, the degree is very considerable and may be well-nigh insurmountable (see Dreyfus [1979], pp. 1–66). Thus, even if mechanisms are not excluded *in principle* from the class of entities to which sensory experience can be ascribed, they may for a long time be excluded *de facto*. Third: even if we find ourselves forced to admit that *in a sense* a highly complex computer can see objects in the range of its sensors, we can find or generate equally troubling and unclear cases from human perceivers. Many brain-lesioned patients manifest partial perception—able to see discrete bits of objects, though failing to recognize gestalts; able to recognize written numbers but not written letters; able to hear musical sounds but nothing else (see Wilkes [1980]). Here too we are tempted to say that such patients see or hear 'in a sense'; we are just as uncertain of how to describe the capacities of such individuals as we are to describe those of computers. Theoretical considerations will determine some answer which sets out in detail what computers, and these brain-lesioned patients, can and cannot do; such considerations will not, as ordinary language does, regard 'seeing' as an either/or, yes/no matter, nor will the question 'Is *A* conscious of this or not?' arise as such for the scientific enterprise. Instead, we should expect to find careful description of what kinds and levels of information processing we find and do not find in the system (human or artefactual) in question.

I turn now to the fourth class of candidate conscious phenomena, that of the propositional attitudes. These, of course, are among the main topics of study for cognitive psychology; whether we call beliefs and their like conscious, preconscious, subconscious, unconscious or simply non-conscious, whether we talk in terms of beliefs, of representations, or of information processing, here is the stuff of which post-behaviourist psychologies are made. I have already suggested that in everyday life (in our common sense psychological descriptions and explanations) we are inconsistent and vague about the applicability of the adjective 'conscious' to the majority of the propositional attitudes. It is thus *prima facie* implausible to suppose that a blurred, shifting conscious/non-conscious dichotomy which is neither exhaustive nor exclusive could hold much promise for systematic study. The fashionable theoretical notion currently dominating the scene is the notion of a representation; crudely, to postulate a representation is to postulate a psychological state which has content. The content of some states is something to which other states have, or do not have, access; this is supposed to hold right from subdoxastic states (for instance, those representing the texture gradients of visual input) up to the most sophisticated and cognitive which perhaps may be identified with something like our familiar beliefs or expectations *etc*. Some want to postulate proposition-type representations, others insist that we need not only these but also image-type representations; this is irrelevant for the purposes of this paper, because in either case the postulated entities are characterized in such a way as to cut across any distinction we may hope to draw between conscious and non-conscious forms of representations. The varieties and various roles of representations, the types of processing to which we must assume they are subject—these are the topics that will constitute the *explananda* for systematic study. As before, the question of consciousness simply drops out; anything that we needed the term 'conscious' to do can be handled more efficiently without it.

6 We are then, I suggest, thrown back on the idea that 'consciousness' and 'conscious' are terms of the vernacular which not only need not but should not figure in the conceptual apparatus of psychology or the neurosciences,

for the concepts required in those theories will want to group phenomena along different, more systematic principles. We can support this argument by looking briefly at the way we use the notion in our everyday language; for by so doing it will be seen that consciousness is not the sort of thing that has a 'nature' appropriate for scientific study.

Qua term of ordinary language, it is wholly unsurprising that 'consciousness' has no precise definition; few such terms do, and this is as true of non-psychological terms in the vernacular as of psychological terms. Moreover, as with all our other everyday mental terms, our ascriptions of consciousness are based on a thoroughly anthropomorphizing prejudice. We 'spread' mentality across infrahuman species to the extent that their appearance and behaviour approximate to ours, and to the extent that we can relate to or interact with them—the extremely stupid koala or a cuddly puppy are anthropomorphized much more than the far more intelligent pig. Now a scientific psychology cannot work from a basis of anthropocentricity; what it says about organisms must hold, or fail to hold, quite independently of what is said about the species *homo sapiens*. Our use of the vernacular, however, is subject to no such objectifying constraints; and the anthropocentricity of everyday mental ascription means that there is little concern with an alleged rightness or wrongness in our attributions of consciousness or any other mental state. We will call a pig, or a bat, conscious if their behaviour and their activities make it seem convenient and appropriate to do so.

It seems—as I have suggested elsewhere (Wilkes [1978], p. 111*ff.*)—that 'consciousness' should be regarded as a second-order concept; 'intelligence', which certainly is so regarded, provides a useful analogy. Nothing is called intelligent if it can perform just one, or a very few, tasks well, even if it does them with extreme efficiency and sophistication; spiders building intricate webs, fish and birds travelling unerringly thousands of miles, chess playing computers, pocket calculators or the *idiots savants* who can juggle numbers with lightning speed—none of these is intelligent.[1] Intelligence is ascribed as a function of the degrees of skill, flexibility and originality with which a number of tasks of diverse sorts can be performed. Put another way: *I.Q.* tests do not test intelligence, but rather test performance at a set of first-order tasks, and intelligence is ascribed to the extent that these are well done. It is thus a second-order predicate in the sense that its ascription rides pick-a-back upon the ascribability of a number of directly testable predicates. That it has a healthy adverbial form ('intelligent*ly*') bears out the claim that it is not a first-order ascription, since first-order ascriptions tend to lack the adverbial cognate.

---

[1] Indeed, Christopher Evans has devised a cross-species *IQ* test, with humans setting the standard at the top with a notional score of 1,000,000, on which, for example, planaria reach scores of tens or hundreds, dogs and cats around 300,000. By this scale the most sophisticated computers in 1978 came out ranking approximately level with the earwig (although it is sobering to realize that only nine years earlier they tallied along with tapeworms—that step took the ancestors of humanity millions of years). See the *Daily Telegraph*, 21 November 1979.

Consciousness is in these respects very like intelligence. The term has a familiar adverbial form; it too is parasitic for its ascription upon the applicability of an assortment of other, more directly ascribable, predicates. In other words, we presuppose a whole slew of psychological ascriptions—to do with perception, motivation, belief and desire, misperception, illusion, recognition *etc.*—when an ascription of consciousness makes sense; conversely, where some set of first-order mental statements are appropriate, then the 'fact' of consciousness follows automatically—there is no further question to ask. Of course, and as we have already noticed in the case of 'see', ascription of first-order predicates can be a highly complex business, quite apart from the need to take account of our anthropocentric perspective. I cannot specify necessary and sufficient conditions for the applicability of 'see', 'hear', 'want', 'afraid' and so forth any more than I can for 'conscious', because I do not believe that there are any such. Nonetheless, the legitimacy of ascribing consciousness will be a function of what other mental predicates we want to ascribe, and the legitimacy of ascribing these is a function of the range of behaviours open to the agent, the range of stimuli to which he can react discriminatingly, and perhaps above all of our attitude to him or it—a robot candidate for consciousness would stand a better chance if it resembled a 'Sleeper' type humanoid than if it were a Dalek.

Consciousness differs from intelligence in that 'intelligence' can be and has been adopted by psychology as an *analysandum*. Could not consciousness equally be so adopted? It will be instructive to see why intelligence is amenable to systematic study, whereas consciousness (I contend) is not. Concepts adopted by a science are almost invariably adapted: given more precise definitions, tailored to fit the needs of the theory. 'Intelligence' in scientific psychology—although not in the vernacular—is operationally defined (indeed, here we find the very paradigm of an operational definition) in terms of success at *I.Q.* tests. The extensions of 'intelligence' in psychology and in the vernacular are roughly equal; those of 'energy' in physics and common sense less so; those of 'spin', again comparing the term in physics and in everyday use, still less close, while 'charm' and 'colour', when applied to quarks, have nothing in common at all with their everyday ancestors. The further the scientific meaning diverges from the vernacular one, the less convincing it is to say that the science in question is analysing or explicating the phenomenon picked out by the ordinary language term. It is obviously true that psychology could, if it wished, adopt and adapt the term 'consciousness' to suit its theoretical needs. But before the outsider can agree that this would indeed be a discussion of *consciousness* (*i.e.*, of what *he* means by 'consciousness') we would need to see just how the science characterizes the notion; and an immediate implication of section 3 above is that no precise characterization could come close to capturing the thoroughly *im*precise and heterogeneous everyday meaning that the term has in the vernacular. The adaptation that the term 'consciousness' would need to undergo before it could be made to cover tidily a systematically-related bunch of behaviours would be so great that a study of *this* 'consciousness' would no more be a study of consciousness as we think of it than the study of the spin of an electron can inform us about the behaviour of whipping-tops.

In sum, then: the fact that an ordinary-language term covers a heretogeneous range of phenomena need not debar it from study (in tidied-up form) by science—'pain' can be and is studied; the fact that a term is a second-order term, like 'intelligence', does not prevent the same treatment; but 'consciousness' is both at once, while compounding its intractability, unlike 'pain' and 'intelligence', by the further fact that its meaning is or seems to be *completely* exhausted by its use: no adequate truth-conditions can be imposed upon ascriptions of consciousness.

7 It is time to turn to the most explicit recent attempt to specify just what it is that a psychology indifferent to consciousness might be missing, as provided by Nagel's phrase, 'What is it like (for an *X*) to be an *X*?' (Nagel [1974]; and also his [1979]). There is, the suggestion runs, something that it is like to be a conscious entity—such as a human, or a bat—and nothing that it is like to be a hamburger or an IBM computer. If an *X* is conscious, then there is something that it is like for that *X* to be an *X*; if not, not. This is the subjective character of experience, essentially connected with a single point of view; something that must be lost by objective science. Moreover, subjective character (as we know from our own case) is highly specific, and it includes an enormous amount of detail, even though we do not possess the vocabulary to describe it adequately; we are compelled to recognize the existence of facts—facts, say, about a bat's subjective experience—which do not consist in the truth of propositions expressible in a human language. 'If we acknowledge that a physical theory of mind must account for the subjective character of experience, we must admit that no presently available conception gives us a clue how this could be done. The problem is unique. If mental processes are indeed physical processes, then there is something it is like, intrinsically, to undergo certain physical processes. What it is for such a thing to be the case remains a mystery' (Nagel [1974], pp. 445–6).

There is little doubt that Nagel's account is the single best attempt to capture the residual *X*-factor that apparently eludes systematic study by the behavioural and brain sciences. The notion of a subjective point of view, ineffable and inexpressible but sharp, complex and known to the subject, explains why the epistemological features discussed above are thought to hold of many conscious phenomena; it explains the perennial temptation to which the private linguist tries to succumb; and it unifies the second, third and fourth of the divisions of consciousness listed above. All the same, I do not think that we yet have a *problem* for psychology and neuroscience; nor do I think there will be much future in Nagel's suggestion that as a first step we should try to supplant orthodox psychology with an objective phenomenology.

One rather blank argument—blank, because there seems to be little that can be done with it—is to object that one does not share Nagel's intuitions. That is, I would deny that I know what it is like to be me, or human, or female; consider Wittgenstein's challenge: 'Can you imagine absolute pitch, if you have not got it?—Can you imagine it *if* you have it?—Can a blind man imagine seeing? Can *I* imagine it?—Can I imagine spontaneously reacting thus and so, if I don't do it? Can I imagine it any better, if I do do it?' (Wittgenstein [1967], p. 49e). (Substituting 'know what . . . is like' for 'imagine' brings out the point clearly.) There is of course a harmless sense in which I know what it is like to be me: I can describe myself to myself and equally to others. But I am aware of nothing that qualifies as *knowledge* behind or beyond that—nothing as significant as Nagel claims. Put another way: a psychiatrist who has remained sane throughout a life devoted to the study of schizophrenia knows more, and better, about what it's like for the schizophrenic to be schizophrenic than does the lay sufferer; put in the form of an objection to Hume, it seems to my intuition that a congenitally deaf scientist who (improbably) had devoted his career to the study of the phenomenology and psychophysiology of auditory perception would know far more than I do about hearing.

These are *just* conflicting intuitions (and perhaps more share Nagel's than share mine); little more can be done than simply to state them. However blank the point, though, the clash of intuition is not worthless; it suggests the metaphilosophical point that intuition may not be a wholly reliable weapon, at least in the arsenal of arguments about consciousness. It is a reasonably well-behaved tool when we are considering possible-world cases involving clear natural kinds; but the further we are from such kinds, the more diverse and variable become the relevant intuitions—and the job of this paper could be described as arguing that consciousness is very *un*like a natural kind. Hence, although the mere affirmation of a conflicting intuition cannot of course overturn Nagel's case, it can weaken it to the extent that it questions the trustworthiness of a premiss: that there is something precise, particular, but ineffable about what it is like for an X to be an X, facts which do not consist in the truth of propositions expressible in any language.

A more direct objection is to query the coherence of the postulation of such puzzling facts. Thus to object would be to marshal some of the familiar arguments against 'the myth of the given', or against the view that there can indeed be knowledge-by-acquaintance which is *not* knowledge by (or under) a description; it challenges the idea that facts can be ineffable even though nonconceptually effed by their possessor. If the Wittgensteinian-cum-Strawsonian arguments against the presuppositions of a private language go through, the subject cannot create *a novo* a subjective phenomenology for himself; and it seems to me that the arguments do go through—at least, I do not see just where Nagel would fault them. A more direct way of putting the point is to object that 'having *A*' need not entail 'knowing (about) *A*'—I can have an inkblot on my face and be the last to know of it. It is true that having *A* does sometimes put me in a privileged epistemological position about it, and this I enjoy *vis-à-vis* any clearly conscious mental states or events. But the knowledge thus possessed is public (or publicisable) enough, so cannot be what Nagel is after; and it is hard sensibly to characterize what he *is* after: knowledge that is *a priori* restricted to acquaintance gives us no way of understanding 'facts' or 'knowledge', still less particular facts and precise knowledge.

Evidently psychology, seeking to transcend the individual, and indeed the human, point of view will not itself say much that is enlightening about subjective points of view; but it has no need to worry about this omission. Just as a subway map often helpfully has an ancillary arrow stating 'YOU ARE HERE', so the notional, complete psychophysiological-physical picture of the world could be supplemented for its users: 'THIS IS YOU', from which the way things are experienced from that position may be inferred. The addition YOU ARE HERE to a map is, evidently, not the cartographer's business, and such maps could not be sold at bookstores; similarly, the addendum THIS IS YOU to a completed psycho-physical account is not a proper part of the scientist's concern. But a fixed map in a subway station on which the station master could stick his arrow, or an objective psycho-physical description to which one could *give* the subjective 'fix'—and both would be heuristics, assisting the user to interpret the maps—could each supply the point of view: from here, Victoria is four stops east, from this perspective, red and green would be indistinguishable (*i.e.*, the perspective of a colourblind man). Psychologists and neuroscientists, like cartographers, must first frame their representations; then, if they want to be helpful, will show us how to interpret and use them—will show us how to find 'the subjective point of view'.

To sum up: sciences don't and needn't explain every phenomenon under every description. No science, for example, explains what it is to be an ornament. Thus there seems no particular reason to suppose that science has, or fails to grapple with, the question of 'what it is like for an *X* to be an *X*'—especially if this is said to be inexpressible and ineffable even within the far richer conceptual resources of the vernacular.

8 I conclude on a more speculative note. I suggest—no more—that just as science can dispense with the concept of consciousness and lose thereby none of its comprehensiveness and explanatory power, so too could ordinary language. Evidently, this is not a recommendation that we should drop the term; few proposals for linguistic reform have any effect. The suggestion is rather than the presupposition that consciousness is an important, or a real, phenomenon should be dropped. The arguments sketched above, although designed primarily for the science, can be readily adapted to show that even in everyday discourse it has little explanatory or descriptive value. Perhaps 'conscious' is best seen as a sort of dummy-term like 'thing', useful for the flexibility that is assured by its lack of specific content; and it is important to remember that the vernacular has countless functions to fulfil *other* than those (description and explanation) that it shares with science—'conscious' may indeed do sterling service in other roles of language. It is description and explanation, and the assumption of realism, that have concerned me; and as a parting note I would point out that the Greeks, who by the fifth century *BC* had a rich, flexible and sophisticated psychological vocabulary, managed quite splendidly without anything approximating to our notion of 'consciousness', nor—a point relevant to section 3 above—did they feel the need to talk about experience*s* (note the plural). One could go further and remark that they had nothing corresponding to 'mind' or 'mental' either—but an attack on the utility of those notions would need to be the subject of another paper.[1]

St Hilda's College, Oxford

REFERENCES

AYER, A. J. [1964]: 'Privacy' in Ayer, A. J., *The Concept of a Person and Other Essays*. London, pp. 52–81.

DENNETT, D. [1978]: 'Toward a Cognitive Theory of Consciousness', in Wade Savage, C. (*ed.*), *Perception and Cognition: Issues in the Foundations of Psychology*. Minnesota Studies in the Philosophy of Science, vol. IX, Minnesota, pp. 201–28.

DREYFUS, H. L. [1979]: *What Computers Can't Do: the Limits of Artificial Intelligence*. New York.

FRANKFURT, H. [1971]: 'Freedom of the Will and the Concept of a Person', *Journal of Philosophy*, **68**, pp. 5–20.

FREUD, S. [1964]: *New Introductory Lectures in Psychoanalysis* (*ed.* and *tr.* Strachey, J.). New York.

HILGARD, E. R. [1977]: *Divided Consciousness: Multiple Controls in Human Thought and Action*. New York.

HOLZMAN, P. S., ROUSEY, C. and SNYDER, C. [1966]: 'On Listening to One's Own Voice: Effects on Psychophysiological Responses and Free Associations', *Journal of Personality and Social Psychology*, **4**, pp. 432–41.

JAMES, W. [1890]: *Principles of Psychology*, vol. II. New York.

JAMES, W. [1912]: *Essays in Radical Empiricism* (Perry, R. B., *ed.*). London.

JOYNT, R. J. [1981]: 'Are Two Heads Better than One?', *The Behavioral and Brain Sciences*, **4**, pp. 108–109.

MELZACK, R. and WALL, P. D. [1970]: 'Psychophysiology of Pain' in Yamamura, H. (*ed.*), *Anaesthesia and Neurophysiology, International Anaesthesiology Clinics*, **8**.

NAGEL, T. [1974]: 'What is it Like to be a Bat?', *Philosophical Review*, **83**, pp. 435–50.

NAGEL, T. [1979]: 'Subjective and Objective', in Nagel, T. *Mortal Questions*. Cambridge, pp. 196–213.

POPE, K. S. and SINGER, J. L. [1978]: 'Regulation of the Stream of Consciousness: Toward a Theory of Ongoing Thought', in Schwartz, G. E. and Shapiro, D. (*eds.*), *Consciousness and Self-Regulation: Advances in Research and Theory*. New York, pp. 101–37.

RORTY, R. [1970*a*]: 'Incorrigibility as the Mark of the Mental', *Journal of Philosophy*, **67**, pp. 399–424.

---

[1] Earlier versions of this paper benefitted from discussions at the Inter-University Centre at Dubrovnik, the Universities of Minnesota, Bradford, Zagreb and Belgrade, and Brasenose College Oxford. I am particularly grateful to Dr Jan Zytkow and Shawn Lockery for their constructive criticisms.

RORTY, R. [1970*b*]: 'Cartesian Epistemology and Changes in Ontology', in Smith, J. E. (*ed.*), *Contemporary American Philosophy*. London, pp. 273–92.

SHOEMAKER, S. [1975]: 'Functionalism and Qualia', *Philosophical Studies*, **27**, pp. 291–315.

WEISKRANTZ, L., WARRINGTON, E. K., SANDERS, M. D. and MARSHALL, J. [1974]: 'Visual Capacity in the Hemianopic Field Following a Restricted Occipital Ablation', *Brain*, **97**, pp. 709–28.

WEISKRANTZ, L. [1980]: 'Varieties of Residual Experience', *Quarterly Journal of Experimental Psychology*, **32**, pp. 365–86.

WILKES, K. V. [1978]: *Physicalism*. London.

WILKES, K. V. [1980]: 'More Brain Lesions', *Philosophy*, **55**, pp. 455–70.

WITTGENSTEIN, L. [1967]: *Zettel* (Anscombe, G. E. M. and von Wright, G. H. *eds*.). Berkeley.

Brit. J. Phil. Sci. 35 (1984), 245–251 *Printed in Great Britain*

Discussion Note

BAYESIANISM AND SUPPORT BY NOVEL FACTS*

1 INTRODUCTION

Gardner in his [1982] assesses the ability of a number of methodological theories to explain the phenomenon, or alleged phenomenon, of the high supportive power, relative to some hypothesis, of any novel facts it predicts. There are, perhaps surprisingly, several definitions of 'novelty' around in the literature, and Gardner himself provides one more. They all have in common, however, that the 'novel' fact is not one which the hypothesis was deliberately constructed in order to explain (*cf.* Zahar [1973], 2.2, Gardner, *op cit.*, p. 11).

One of the methodological theories cited by Gardner is Bayesianism, and it is charged by him with failure to explain how *any* fact 'novel' in this sense but known at the time a hypothesis was formulated can support that hypothesis. My aims in this note are twofold: they are, first, to correct the elementary misunderstanding of the Bayesian theory responsible for that verdict and, second, to show how a Bayesian analysis provides a more accurate criterion than that of Zahar and others (according to which it is whether they are 'novel' or not, in the minimal sense above) for determining which *known* facts support subsequently-formulated hypotheses and which do not.

2 BAYESIANISM

The Bayesian's fundamental explanatory concept is a conditional probability function $P(A|B)$ relativised to a stock K of background information and some (rather idealised) cognitive agent. $P(A|B)$ measures, where K is for the time being assumed not to contain B, what this individual's degree of belief would be in A were he to come to know B.[1] The support, relative to K, that he would regard evidence E as giving a hypothesis H in the context of K is then measured by $P(H|E)-P(H)$.

But suppose K already includes E. $P(H|E)$ relative to such a K is then equal to $P(H)$ if $H \vdash E$; and were these two quantities substituted into the support measure the value of the latter would be zero. Obviously this is undesirable. On the other hand, it is equally clearly inappropriate to attempt to measure $P(H|E)$ and $P(H)$ relative to the state of knowledge just before E was discovered (especially if this were some considerable time before H is first contemplated). Gardner concludes from such considerations that the 'Bayesian assumptions' which allegedly imply that the support E gives H at time t is equal to $P_t(H|E)-P_t(E)$, therefore entail that evidence already known at t gives no support at t (*op. cit.*, p. 13). The objection is not new: it originates with Glymour ([1980], ch. 3 to which Gardner refers). I shall comment on Glymour's own contribution presently.

---

* I am very grateful to Peter Urbach for his many helpful criticisms of earlier versions of this paper.

[1] The standard argument for supposing that conditional degrees of belief are measured by a conditionalised probability function is due to de Finetti [1937], pp. 108–10.

But there are no 'Bayesian assumptions' implying such a way of reckoning support when E is already known, as any careful reading of what Bayesians themselves have to say would have revealed. In particular, there is nothing in the account above to suggest that knowing that E is true entails that the support measure must incorporate a P function relativised to information containing E. Indeed, there is every reason why it should not. For the Bayesian assesses the contemporary support E gives H by how much the agent would change his odds on H *were he now* to come to know E. This means that in the context of K he will compute the difference between what his degree of belief would be in H assuming (counterfactually if he already knows E) that his state of knowledge were $K-\{E\}$, and what his degree of belief in H would be were he, knowing $K-\{E\}$, to come to know E. It follows that the support he regards E as giving H is equal to $P(H|E)-P(H)$, both probabilities being relativised to $K-\{E\}$[1] whether he knows E or not. In other words, the theory is explicitly a theory of ***dispositional*** properties of the agent's belief-structure, though the implicit conditions may not be realised in practice.

Glymour considers something like this account[2] but only as a sort of *ad hoc* modification of what he calls 'Bayesian kinematics' (*op. cit.*, p. 86), according to which the support given to H by E at time t is alleged to be equal to the conditional probability of H on E at t minus the probability of H at t, where the probability function is relativised to the *total* available information available at t. In other words, Glymour's 'Bayesian kinematics' are Gardner's 'Bayesian assumptions'. But 'Bayesian kinematics' are not part of Bayesian theory. The latter is quite immune to the charge of making either no or alternatively outrageous assessments of support by known facts (though it is not immune to being misapprehended); as we shall shortly observe, the criteria yielded by a Bayesian analysis in this context are eminently reasonable.

---

[1] $\{E\}$ will be taken to include everything in K dependent on E.

[2] And considers it applicable only in the evaluation of simple statistical models (p. 88), but not to 'actual historical cases'. His principal argument appears to be the difficulty of grasping the logical relations between hypotheses and observation reports. But this is an objection not, as he seems to think, to positing a value of $P(E)$ relative to the sort of counterfactual background information discussed above, but to the idealisation involved in requiring the P-function to respect logical relations.

3 EVALUATING SUPPORT BY KNOWN FACTS

However, the Bayesian account does not seem to explain why those facts predicted by a hypothesis but also used in its construction do not support it, as Zahar [1973], p. 102 and Worrall [1978], p. 48 insist they do not. But this 'phenomenon' is better left unexplained, for it is entirely bogus. Indeed it is only too easy to find counterexamples. Thus, suppose we do not know the proportion p of black balls in an urn containing just black and white balls. In a long sequence of $m+n$ drawings with replacement we find the frequency of black balls to be $m/(m+n)$. As a result we propose the hypothesis that the ratio we seek is $m/(m+n)\pm\varepsilon$ for suitably small ε. If we regard statistical estimation theory as a reasonably scientific procedure, as I think most of us do, then this hypothesis is very well supported by the evidence from which it was obtained—obtained indeed as the best estimate of the ratio.

This is undisputed ground whether one is a Bayesian or not: it is standard statistical methodology. If one is a Bayesian one will justify the methodology as follows: we know that the appropriate model is very likely to be that of a sequence of Bernoulli trials with unknown parameter p, where $0 \leqslant p \leqslant 1$. With a uniform distribution and sufficiently large values of $m+n$, $P(m/(m+n)-\varepsilon < p < m/(m+n)+\varepsilon|E)$ is close to 1, while $P(m/(m+n)-\varepsilon < p < m/(m+n)+\varepsilon) = 2\varepsilon$. Hence for small ε and sufficiently large $m+n$ the support given the hypothesis by E is nearly maximal. But it is to be emphasised that this is merely a Bayesian explanation of a verdict which is fairly unanimously delivered, namely that in appropriate circumstances evidence used directly in the construction of hypotheses most certainly does support those hypotheses.

What are 'appropriate circumstances'? At least the following two seem necessary: first, the prior plausibility of the model itself must be reasonably high (it is, for the sort of example we gave); and second, the adjusted hypothesis *i.e.* the model with its parameters calculated from the data (for example, by a confidence interval as above) has actually to yield an explanation of the data. Both these criteria are satisfied in the example. We can illustrate the effects of varying prior probability and explanatory power as follows. Let H' be the hypothesis in the example with its parameter p estimated in the same way from the sample. By Bayes's Theorem

$$P(H'|E)-P(H') = P(H')\left[\frac{P(E|H')}{P(E)}-1\right].$$

Now $P(H') = 2\varepsilon.\,P(H)$,[1] where H is the hypothesis merely that the model is a Bernoullian process. Hence

$$P(H'|E)-P(H') = 2\varepsilon\,.\,P(H)\left[\frac{P(E|H')}{P(E)}-1\right]-(1)$$

---

[1] $P(H') = P(H'|H)\,P(H)+P(H'|\bar{H})\,P(\bar{H}) = 2\varepsilon P(H)+0$.

It was argued by Good in his [1968] and recently by Howson [1983] that $P(E|H')/P(E)$, or an increasing function of it like the logarithm, is a measure of the degree to which H' explains E. As an illustration, suppose that in the example above each black ball is followed by a white ball followed by a black ball *ad infinitum*. It seems wrong completely to rule out, *a priori*, deterministic models, since we do not know with certainty that they might not apply.[1] Now the number of finite rules, computable and not, which generate distinct binomial sequences is denumerable, and so each may be assigned a finite prior probability (and the simpler the rule, say, the higher the probability). $P(E) \geqslant P(E|H_1)P(H_1)$ where H_1 is the rule of alternate black and white appropriate to E, and so $P(E)$ will for this E be equal to a finite number. Since the likelihood term $P(E|H')$ tends to 0, so too does $P(E|H')/P(E)$. If on the other hand the sequence of individual outcomes is 'random', in the sense of refuting at some value of $m+n$ each rule-hypothesis, leaving in the limit an independence model as the only plausible alternative, then $P(E)$ tends to 0 and the ratio $P(E|H')/(E)$ tends to a limit proportional to ε^{-1}. If this sort of measure of explanatory power is accepted, then (1) tells us that the support of H' by E is an increasing function of both the prior probability of the *model* H and the extent to which its derived H' explains the sample E. At any rate, we can infer that if $P(E|H')/P(E) > 1$ then E gives positive support to H', to degree depending on the prior plausibility of the Bernoullian model; and this is wholly consistent with what our intuitions tell us is the case. Indeed, the intuitive force of such simple counterexamples to Zahar's and Worrall's thesis is, I think, irresistible. The Bayesian analysis merely confirms it and indicates rather accurately in which circumstances facts like those recorded in E support hypotheses like H' constructed from them. To sum up the discussion so far: it is not a question of *whether* facts used in the construction of a theory support it, but *when*.

The analysis given above is by no means restricted in application to statistics, but is quite general, with the proviso that in the deterministic case $P(E|H')$ will (ideally) be equal to 1. Suppose H is a general functional hypothesis with some of its parameters unspecified, which provides no grounds for supposing that any one set of values of those parameters is more likely than another, within certain broad limits. E records as many simultaneous readings of the variables as are necessary to fix the parameters of H, which done we get H'. Intuitively the support given H' by E should certainly be non-zero, and we would probably reason as follows. The odds on H' justified by E are plausibly the same as our prior odds on H, since H' is no more and no less endorsed by background information plus E than H by background information alone. Moreover the prior odds on H' should be considerably, possibly orders of magnitude, lower than those on H, assuming that we have no independent grounds for expecting with anything approaching certainty that the magnitudes reported in E take the value they are known to have. It follows that the difference between posterior and prior odds on H' should certainly be positive.

---

[1] It seems reasonable to suppose that *e.g.* human agency *may* be responsible for the observed outcomes.

The Bayesian reasoning is pretty much the same. Our first assumption is suitably expressed by the condition of independence between H and E, *i.e.*

$$P(E \wedge H) = P(H)P(E).$$

Also $H' \wedge E \Leftrightarrow H' \Leftrightarrow H \wedge E$, modulo background information including the initial conditions for E. Hence

$$P(H'|E) = \frac{P(H' \wedge E)}{P(E)} = \frac{P(H')}{P(E)} = \frac{P(H \wedge E)}{P(E)} = \frac{P(H)P(E)}{P(E)} = P(H).$$

Therefore

$$P(H'|E) - P(H') = P(H)[1 - P(E)] > 0 \quad \text{if} \quad P(E) < 1 \tag{2}$$

which corroborates the informal reasoning above.

It follows that according to a Bayesian analysis the support for H' from E is determined by the two quantities $P(H)$ and $P(E)$, and that

$$0 \leqslant \text{Support}(H', E) \leqslant P(H)$$

so that $P(H)$ is an upper bound and the support approaches 0 as $P(E)$ approaches 1. Now suppose that for some other hypothesis H_0 with positive prior probability, $H_0 \vdash E$. Then $P(E) \geqslant P(H_0)$. Hence if $P(H_0)$ is not negligible and $P(H)$ is small the support of H' by E will be very inconsiderable. For a rather striking illustration of such a situation, let H' be Newtonian gravitational theory H together with values of a postulated solar oblateness which are determined by the observed perihelion advance of Mercury reported by E, and let H_0 be General Relativity. Given present background knowledge incorporating all the experimental results strongly confirming Special Relativity, $P(H)$ is small, and since H_0 predicts, to within the bounds of experimental error, the value recorded in E, equation (2) predicts low support for H'. This result, independently of measurement of the sun's departure from sphericity, seems to concur well with the standard view (though the solar oblateness hypothesis, adjusted to yield the observed perihelion shift of Mercury, does not fit those of the other planets as well as Einstein's theory[1]).

Zahar notes (*op. cit.*, p. 257) that Einstein's prediction of the magnitude of Mercury's perihelion advance, in contrast to its rather *ad hoc* incorporation into the classical framework, early on told in favour of the former. Now it follows immediately from the equations above that H_0 and H' enjoy equal support from E if and only if $P(H_0) = P(H)P(E)$. If H_0 and H, therefore, are rival general theories, to the extent that they have approximately equal probabilities, and if the only ground for any expectation of E is its prediction by H_0, so that $P(E) \ll 1$, then the support of H_0 by E is greater than that of H': H''s support relative to H_0 is discounted precisely by the probability of E. Thus the fact that Zahar draws our attention to, namely the superior level of support which people thought Einstein's explanation of Mercury's perihelion advance received in 1915, as opposed to some introduction of adjustable parameters into classical gravitation theory, is explained quite simply, *consistently* with the denial that hypotheses like H' necessarily fail to be supported by the data from which they were constructed.[1]

---

[1] Adler, Bazin and Schiffer [1965], pp. 205, 206.

In examples like this we must be careful to keep two considerations distinct: (i) the support by E of the general model H, and (ii) the support by E of the derivative hypothesis H'. In general E will not support H, and it is this fact which together with the conflation of (i) and (ii) seems to have been responsible for the belief that facts used in the construction of a theory do not support it. In Worrall's case the conflation is apparent in his assertion (*op. cit.*, p. 48) that 'Mercury's perihelion [advance] is not regarded as supporting classical theory' when it is predicted by some *versions* of that theory constructed precisely for this purpose. This is true, but it does not follow that those versions do not thereby acquire some support. In particular, his conclusion that 'of the empirically accepted logical consequences of a theory those, and only those, used in the construction of the theory fail to count in its support' (*ibid.*) is false.

4 CONCLUSION

The Bayesian treatment of known facts is misunderstood by Gardner and Glymour, and the known-ness or otherwise of evidence causes it no problems. Indeed, we have just seen that the Bayesian theory gives a simple and elegant account of why a theory which provides strongly *ad hoc* explanations of data, in the sense that it has a sufficient number of adjustable parameters, does not receive as much support from those data as a rival theory which directly explains the same data. This central methodological fact is explained by Zahar and Worrall in a way which commits them to the denial that theories are ever supported to any extent by data which they explain in such an *ad hoc* way. Such a view, for the reasons advanced earlier, makes nonsense of quite basic and eminently reasonable scientific appraisals. The Bayesian explanation of why this sort of *ad hoc*ness depreciates the support given by data so explained does not, however, entail any such unacceptable corollary; on the contrary, it indicates very accurately the sorts of circumstances in which positive support from such data might be forthcoming.

COLIN HOWSON
London School of Economics

---

[1] Note that though in any one case the support can be made zero by choosing a different origin, this cannot be done uniformly. The phenomenon of positive support is not merely the result of a judicious choice of scale.

REFERENCES

ADLER, R., BAZIN, M. and SCHIFFER, M. [1965]: *Introduction to General Relativity.*

DE FINETTI, B. [1937]: 'Foresight, its logical laws, its subjective sources', reprinted in *Studies in Subjective Probability*, edited by H. E. Kyburg and H. E. Smokler.

GARDNER, M. R. [1982]: 'Predicting Novel Facts', *British Journal for the Philosophy of Science*, **33**, pp. 1–15.

GOOD, I. J. [1968]: 'Corroboration, Explanation, Evolving Probability, Simplicity and a Sharpened Razor', *British Journal for the Philosophy of Science*, **19**, pp. 123–43.

GLYMOUR, C. [1980]: *Theory and Evidence.*

HOWSON, C. [1983]: 'Statistical Explanation and Statistical Support', *Erkenntnis*, July 1983.

WORRALL, J. [1978]: 'The ways in which the Methodology of Scientific Research Programmes Improve on Popper's Methodology' in G. Radnitzky and G. Anderson (*eds.*), *Progress and Rationality in Science*, pp. 45–70.

ZAHAR, E. [1973]: 'Why did Einstein's Programme supersede Lorentz's?', *British Journal for the Philosophy of Science*, **24**, pp. 95–123, 223–62.

Brit. J. Phil. Sci. 35 (1984), 253–292 *Printed in Great Britain*

Review Articles

POPPER, PROPENSITIES, AND QUANTUM THEORY*

1 *Popper's new thought experiment*
2 *Popper's propensity interpretation*

In these volumes the reader will find an even better Popper—the same lucidity and economy of prose, but unencumbered by the rather strict demarcation criterion of *The Logic of Scientific Discovery*; and hence free to speculate on matters metaphysical. Indeed it is this particular change—the embracing of metaphysics—which strikes one as the most significant change from the Popper of *L.S.D.*:

> But if my dream is metaphysical, what is the use of it? Is there anything in it beyond, perhaps, an emotional satisfaction? Is it not utterly different from a scientific hypothesis—one in which we are mainly interested *because of its implicit claim to be considered tentatively as true?*
>
> I no longer think, as I once did, that there is a difference between science and metaphysics regarding this most important point ([1983], p. 199).

However Popper goes on to say:

> . . . *as long as a metaphysical theory can be rationally criticised*, I should be inclined to take seriously its implicit claim to be considered, tentatively, as true.

; and [1983], page 200:

> Any critical discussion of it will consist, in the main, in considering how well it solves its problems; how much better it does so than various competing theories whether the solution is simple; how fruitful it is. . . .

Not so much of a change after all, but a change nonetheless; and it is in the spirit of this change that I shall here address myself to a critical evaluation of Popper's metaphysics.

1 POPPER'S NEW THOUGHT EXPERIMENT

Popper proposes a new metaphysics for Quantum Theory (*QT*) in [1982b]. A central theme in his argument for this new metaphysics is a repudiation of the 'subjectivism' (and more generally of the 'anti-realism') of some versions of the Copenhagen Interpretation (*CI*)—particularly as it is represented in the early works of Heisenberg. For example, Popper explicity sees himself in opposition to Heisenberg's view that:

---

* Review of K. R. POPPER [1983]: *Realism and the Aim of Science*, and [1982]: *The Open Universe*, and *Quantum Theory and the Schism in Physics*, from the Postscript to the Logic of Scientific Discovery, edited by W. W. Bartley, III. Hutchinson.

> objective reality has evaporated and that quantum mechanics does not represent particles, but rather our knowledge, our observations, or our consciousness of particles.[1]

In the preface to [1982b] (pp. 27–30), Popper presents an entirely new thought experiment (different from his own much earlier prototype for the Einstein Podolski Rosen experiment), which purports to provide a crucial test for the *CI* (although not for *QT* itself); and which he is 'inclined to predict' will decide against the subjectivist. *CI.* Since new crucial tests in the area of *QT* have an intrinsic interest, and since the claims for this one are so far-reaching (a promise of 'subjectivism' empirically refuted), I shall spend some time discussing it in this section. In section 2, I shall then move on to a more general assessment of Popper's new metaphysics—and in particular of his propensity interpretation for probabilities, which he regards as vindicated by its role in a coherent metaphysical framework for *QT* (and in particular by its role in an explanation of the 'double-slit experiment'). Note that in the first section I shall, for argument's sake, follow Popper, Einstein, *et al.* in assigning the state-descriptions of *QT* to ensembles of systems (or, in more Popperian terms, to a replicatable experimental set-up) rather than to particular systems. I shall question this later, in section 2.

Popper's experiment is simply described as a hybrid of an Einstein Podolski Rosen (*EPR*) set-up and a single slit diffraction experiment. Consider an ensemble of pairs of particles, prepared in an *EPR* state, so that there is probability 1 of agreement between the measured values for the y-coordinates of the members of any given pair.[2] (Popper's "realism" means that he would here talk of actual values rather than measured values; but this difference has no immediate significance in what follows, and I shall therefore ignore it.) Now suppose that one of each pair of particles—call it the 'A-particle'—moves off in the positive x-direction, whereas its partner—the 'B-particle'—moves off in the opposite direction. In the positive x-direction is located an opaque screen with a slit in it—'slit A'—oriented along the y-direction; and in the negative x-direction is located a similar screen with slit b in it. Behind the slits are located various counters to register the angles at which the various particles which pass through the slits scatter (Figure 2, [1982b], p. 28).

I shall now quote from Popper ([1982b], pp. 28–32). I make appropriate substitutions to accord with my different notation):

> Now we make the slit A very small and the slit B very large. . . .

Doing this means that we make a direct and fairly precise measurement of y

---

[1] Quoted from Heisenberg by Popper [1982b], p. 35.

[2] The *EPR* state further requires agreement between the measured values for the y-component of the momentum of one particle and minus the y-component of the momentum of its partner. But Popper's arguments do not make use of this extra requirement.

(the y-coordinate) for those A-particles which pass through slit A, with an accuracy given by the width of slit A. Because of the *EPR* correlation, this means that we have also made an indirect and equally accurate measurement of y for the corresponding B-particles; *i.e.* we can say that their measured values for y are also within the interval occupied by the slit A. Or, as Popper puts it:

> According the *EPR* argument we have measured the y-position coordinate for both particles . . . with the precision of the width of the slit A.

And now comes Popper's point:

> We thus obtain fairly precise '*knowledge*' about $y(B)$—we have 'measured' it indirectly. And since it is, according to the Copenhagen interpretation, our *knowledge* which is described by the theory—and especially by the Heisenberg relations—we should expect that the momentum . . . of the beam that passes through slit B scatters as much as that of the beam that passes through slit A, even though the slit A is much narrower. . . . If the Copenhagen interpretation is correct, then such counters on the far side of slit B that are indicative of a wide scatter . . . should now count coincidences; counters that did not count any particles before the slit A was narrowed. . . . To sum up: if the Copenhagen interpretation is correct, then any increase in the precision of our *mere knowledge* of the position of the particles going through slit B should increase their scatter.

The last scatter referred to here by Popper is ambiguous between the initial scatter of measured values for p_y over the B-particles and the resultant spatial scatter of the B-particles along the y-axis after passage through B. But it can be shown that these two scatters are in fact proportional to each other under special circumstances, and hence the ambiguity can be rendered harmless.[1] In order to avoid the complexities of justifying this proportionality however, I shall here take 'scatter' to refer to the former scatter (of measured p_y values); and hence replace the counters behind the slit B in Popper's experiment by a single p_y measuring apparatus. I note also that the slit B has no role to play in Popper's argument; and so I remove it from the set-up.

My simplified version of Popper's experiment now looks like this. There is an ensemble of particle-pairs (A_1, B_1), (A_2, B_2) . . . , prepared in a pure state for which the y-coordinates of the A-particles are correlated with the y-coordinates of the corresponding B-particles; and the A-particles move in the positive-x direction while the B-particles move in the negative-x direction. Now insert a wide slit into the paths of the A-particles. Let w_1 be the width of the slit, and $S_1(p_y)$ be the scatter of the measured p_y values over those B-particles for which the corresponding A-particles pass through the slit. (We estimate this scatter from a finite sample of the sub-ensemble.) Now repeat this experiment for a different ensemble of A and B particles but still in the same correlated state, and with a slit of narrower width $w_2 (w_2 < w_1)$. Let $S_2(p_y)$ be the resultant scatter. Popper claims that the subjectivist version of the *CI* entails that $S_2(p_y) > S_1(p_y)$. *I.e.* the subjectivist *CI* is committed to saying that a mere increase of knowledge in the location of the *B*-particles (by restricting the corresponding *A*-particles to those passing through the narrow slit) causes an objective increase in the scatter of measured values for momentum. And it is Popper, *qua* realist (and hence anti-subjectivist) who opposes this; and in particular predicts that $S_2(p_y) \leqslant S_1(p_y)$. Indeed his argument would seem to commit him to asserting that $S_2(p_y) = S_1(p_y)$.

---

[1] *E.g.* Ehrenfest's theorem (p. 26, Schiff [1955]) shows that on average the momentum of a particle of given mass determines its rate of change of position (in a given direction). Thus a scatter of momentum is reflected eventually in a scatter of spatial location.

However, on page 29 of [1982b], we see that Popper has a fall-back position prepared, from which to save realism, if the crucial test should go against him (*i.e.* if it should turn out that $S_2(p_y) > S_1(p_y)$). This fall-back position is to admit 'action-at-a-distance', *i.e.* admit that the filtration of the *A*-particles (by passing them through the slit) does affect the correlated *B*-particles, despite the spatial isolation of the *B*-particles from the *A*-particles at the time when filtration occurs. Thus, although subjectivism may well be under test in Popper's experiment, realism is not. A positive result for subjectivism can be reinterpreted by a realist as merely (!) action-at-a-distance.

The key to demolishing Popper's argument is to examine in detail the reasons he has for asserting that a subjectivist reading of the *CI* entails that $S_2(p_y) > S_1(p_y)$. For argument's sake I shall initially grant Popper a technical point which he needs, namely that the initial collection of *B*-particles for which the corresponding *A*-particles pass through the slit w_1, is an ensemble for which the quantum-theoretical uncertainties are minimised (later I shall question this point); *i.e.* we have

(a) $S_1(p_y) \text{x} S_1(y) = h$ (where 'h' is short for '$h/2\pi$')

instead of the more generally valid $S_1(p_y) \text{x} S_1(y) \geqslant h$, where $S_1(y)$ is the objective scatter of y-coordinates over that same collection of *B*-particles. We also of course have, quite generally, that

(b) $S_2(p_y) \text{x} S_2(y) \geqslant h$

Now a subjectivist reading of CI requires that

(c) $S_1(y) = w_1$ and $S_2(y) = w_2$.

This is because w_1 is the degree of precision of our knowledge about the y-coordinates of those *B*-particles for which the corresponding *A*-particles pass through the slit of width w_1; and 'subjectivism' then equates this degree of precision in our knowledge with the objective scatter (and similarly for w_2). Thus we see that

$$S_2(p_y) \geqslant \frac{h}{S_2(y)} = \frac{h}{w_2} > \frac{h}{w_1} = \frac{h}{S_1(y)} = S_1(p_y),$$

i.e. $S_2(p_y) > S_1(p_y)$ as required. As indicated above it is the premise (c) of this derivation which Popper questions; and it is the falsity of (c) which Popper uses as an explanation for the failure of '$S_2(p_y) > S_1(p_y)$'—a failure which he predicts in opposition to a subjectivist reading of the *CI*.

But Popper's argument here errs in laying the premise (c) at the door of subjectivism. True, a subjectivist reading of *CI* does imply (c); but (c) can also be justified on far less controversial grounds—from within the central formalism of *QT*, independent of any particular interpretation of *QT*. *I.e.* we simply point out that the measured values for the y-coordinates of the particles within the initial sub-ensemble of A-particles are evenly spread over the space-interval occupied by the slit.[1] Then those particles which pass through the slit will have a scatter for the measured values of their y-coordinates which is precisely w_1; and hence so too will the corresponding B-particles. *I.e.* $S_1(y) = w_1$. And similarly we can justify $S_2(y) = w_2$. This justification of (c) is quite independent of the *CI*—subjectivist or otherwise. Indeed all we need in the justification is the central part of *QT* (the 'Born statistical interpretation') which takes us from the state-vector of a system to the 'observable' probabilities of measuring physical quantities to have particular values. Thus in denying that $S_2(p_y) > S_1(p_y)$, Popper is putting forward a substantial variant of *QT*, and not just a realist interpretation of it. So what is being tested by Popper's thought experiment is not after all the *CI*; rather it is a crucial test between *QT* itself and Popper's variant of it. I predict therefore (in opposition to Popper) that were it possible to perform Popper's experiment, then we would find that $S_2(p_y) > S_1(p_y)$; *QT* would be vindicated and Popper refuted.

In claiming this, of course, I must come to terms with Popper's claim that a realist interpretation of the result that $S_2(p_y) > S_1(p_y)$ indicates action-at-a-distance.[2] Indeed if we follow Popper's line of argument, then even on the *CI* this result seems to imply action-at-a-distance, *i.e.* the change of knowledge of the B-particles entailed by narrowing the distant slit which constricts the correlated A-particles results in a change of scatter of the B-particles.[3] Is Popper right in his claim that action-at-a-distance is indicated if $S_2(p_y) > S_1(p_y)$? In my view, the error in making this claim stems from adopting too loose an interpretation of the term 'scatter'. It can perhaps be argued that action-at-a-distance is indicated in the hypothetical situation where certain measured values for $p_y(B)$ are possible when the A slit is narrow which are not possible when the A slit is wider. But this hypothetical situation is not the same as the situation that $S_2(p_y) > S_1(p_y)$ with which we are confronted here (although Popper says it is). We can see this once we keep in mind the technical sense which is assigned to 'scatter' from within QT (See p. 16, equation (1) of Heisenberg [1930], for example, namely as a 'standard deviation'. For example, consider an ensemble of systems for which a random variable Q has the value 1 with probability 1/8, the value -1 with probability 1/8, and the value 0 with probability 3/4. Then the standard deviation of Q values (the root mean square deviation) is $(2/8)^{1/2} = 1/2$. But now consider that sub-ensemble of this initial ensemble which consists of just those systems which have the values 1 or -1 for Q. The standard deviation of Q values over this sub-ensemble is $(2/2)^{1/2} = 1$. Thus a simple selection process, without any modification of the elements selected, may increase the standard deviation of a collection of elements. In particular we see that action-at-a-distance is not necessary (nor may even be indicated) in order to explain why narrowing the slit increases the standard deviation of the measured momentum values. *I.e.* it can be argued that by narrowing the slit we simply select out fewer of the same elements that would have been selected had the slit been broader; and such a selection process may be quite sufficient (as we have just seen) to account for an increase in 'scatter', *qua* standard deviation. Thus we can accommodate the result that $S_2(p_y) > S_1(p_y)$ within a quite realist version of QT (or Popper's subjectivist version of CI for that matter) without appealing to action-at-a-distance (contrary to Popper's claim).

---

[1] Indeed it is precisely this assumption which is made by assuming that the particles are prepared in an *EPR* state; *i.e.* it can then be shown that there is equi-probability of measuring the y-coordinate of the A-particles in any spatial interval of given width.

[2] See [1982b], p. 29. Popper does express some reservations in making this claim, but it is hard to see how he can go this given his other views. *I.e.* it would seem to follow from his views that if $S_2(p_y) > S_1(p_y)$ then action-at-a-distance has occurred—see below.

[3] See Popper's comments to similar effect on p. 16 [1982b].

In fairness to Popper however, the following remark must be added. For one interpretation of QT—that version of the CI espoused by the later Heisenberg—the scatter in measured values for some physical quantity Q over an ensemble of systems in some (pure) state, reflects an underlying objective indeterminancy in the value of Q for *each* system in the ensemble. In that case it seems that the change in scatter of the measured p_y-values over the relevant collection of B-particles, does reflect an objective change for each B-particle; and hence we get action-at-a-distance. My point holds nevertheless that for neither the subjective version of CI, nor the 'realist interpretation' of QT (such as Popper's), is action-at-a-distance at all indicated by the result that $S_2(p_y) > S_1(p_y)$. I return to this point at the end of this section.

My final criticism of Popper is in some ways the most definitive, but it is also the easiest to put aside. The criticism is simply that the technical point which we granted Popper above cannot after all be granted, and that this has disastrous consequences for Popper's whole experiment. Nevertheless, as we shall see, by doing a 'Bohm–Aharanov' to Popper's experiment we can avoid this particular criticism. Popper's argument relies on assuming that for his initial collection of B-particles (with the slit at its wider setting) we have not just

$$S_1(p_y) \mathrm{x} S_1(y) \geqslant h$$

but also

(a) $$S_1(p_y) \mathrm{x} S_1(y) = h$$

But clearly this is not so! *I.e.* the initial ensemble of all the A-particles is easily seen to be in a mixture of y-coordinate eigenstates[1]—indeed this is easily shown to follow just from there being a correlation between the measured values for the y-coordinates of the A- and B-particles. And if we pass such a mixture through a slit oriented along the y-direction, then clearly those particles which pass through the slit will still be in a mixture of y-coordinate eigenstates (but with a narrower dispersion than the initial ensemble). And hence so too will the corresponding B-particles. But for a collection of particles in a y-coordinate eigenstate—or a mixture of such—the scatter of measured momentum values is easily seen to be infinite. Thus not only does (a) fail to hold, but we have

$$S_1(p_y) \mathrm{x} S_1(y) = \infty.$$

Moreover this failure of (a) is such as to render Popper's thought experiment impotent. *I.e.* if $S_1(p_y) = \infty$ then it is impossible to decide between $S_2(p_y) > S_1(p_y)$ and $S_2(p_y) = S_1(p_y)$.

This technical objection to Popper can be avoided however, by a reformulation of Popper's experiment along the lines that Bohm and Aharanov reformulated the *EPR* experiment [1957]. So this technical objection to Popper proves ultimately to be beside the point. I shall present this reformulation here briefly, in the Popperian spirit of minimising technical exposition. The presentation will be detailed enough however, to show that the other criticisms of Popper which I made above still apply. Suppose that the ensemble of particle pairs (A_1, B_1), (A_2, B_2), . . . is prepared in a pure state

$$\sum_i c_i f_i(A) \mathrm{x} f_i(B),$$

where $f_i(A)$ is the eigenvector of the non-degenerate physical quantity $Q(A)$ (for the A-particles) for eigenvalue q_i; and similarly for $f_i(B)$. We suppose that the A-particles and B-particles are spatially well separated from each other. Now consider two quite different experimental situations. In the first situation we estimate the probability $\mathrm{Prob}_1(P(B), j)$ for a physical quantity $P(B)$ to be measured to have the value p_j over the whole ensemble. *QT* tells us that

$$\text{(d)} \qquad \mathrm{Prob}_1(P(B), j) = \sum_i |c_i|^2 |(f_i(B), g_j(B))|^2,$$

where $g_j(B)$ is the eigenvector of $P(B)$ for eigenvalue p_j. In the second situation we firstly measure $Q(A)$ for the A-particles, and then consider the ensemble of B-particles for which the corresponding A-particles exhibit the value q_i for $Q(A)$. For this second ensemble we measure the probability

---

[1] This follows from the von Neumann rule for deriving the density operator of a sub-system from the density operator of the total system, as applied to the *EPR* state for the initial ensemble of all pairs of A-particles and B-particles.

$\mathrm{Prob}_2(P(B), j)$ of $P(B)$ being measured to have the value p_j. Clearly $\mathrm{Prob}_2(P(B), j) = \mathrm{Prob}(Q(A), i/P(B), j)$, where the latter probability is the conditional probability of $Q(A)$ being measured value q_i given that the measured value of $P(B)$ is p_j. *QT* tells us that this probability is

$$\text{(e)} \qquad \mathrm{Prob}_2(P(B), j) = \frac{|c_i|^2 |(f_i(b), g_j(B))|^2}{\sum_i |c_i|^2 |(f_i(B), g_j(B))|^2}$$

But if we followed Popper's line of reasoning above then we would have to conclude that the scatter of measured values for $P(B)$ in the second situation could not exceed the scatter in the first situation (while holding in reserve that if we were wrong, then this would indicate action-at-a-distance). And it is easy to demonstrate that this conclusion is inconsistent with *QT*, *i.e.* that the scatter, *qua* standard deviation, of the distribution defined by (e) above must, in certain cases, exceed that defined by (d)[1] (without involving any action-at-a-distance effects). So Popper is in conflict with a central part of *QT*, and not just the *CI*.

It is important to note that in coming to this last conclusion one is not committed to action-at-a-distance. *I.e.* for reasons already given, both a subjectivist and a realist *CI* can accommodate the result that the scatter for distribution (e) can exceed that for (d), without invoking action-at-a-distance. Moreover, even the objectivist *CI* (of the later Heisenberg), which I discussed above, can make such an accommodation, by pointing out that the scatter over possible values for $P(B)$ is *not* just a reflection of an underlying objective indeterminancy in $P(B)$. It is also a result of the B-particles being in a *mixture* of $Q(B)$-eigenstates; so that, whether or not a measurement of $Q(B)$ takes place, each B-particle can be taken as showing an objective indeterminancy for $P(B)$ characteristic of one of the $Q(B)$-eigenstates. So again, no action-at-a-distance is indicated.

This completes my discussion of Popper's thought experiment. We see that, contrary to Popper's intention, his views actually contradict a central part of *QT*, and not just the subjectivist *CI*. Moreover, we have seen that if the results of the thought experiment go against Popper's views (and in favour of *QT*), then there is (contrary to Popper's claim) *no* indication of action-at-a-distance. Finally I note that my reformulation of Popper's thought experiment shows clearly that the situation hinted at by Popper—of there being a possible value for $P(B)$ in the second situation which is not possible in the first situation—cannot take place consistently with *QT*, or at least this is so if we take a 'possible value' to be a value with non-zero probability. This is because it is impossible for the right hand side of (e) to be non-zero without the right hand side of (d) being non-zero too. In this way a crucial test of Popper's views is made possible. And if it is standard QT which survives—as one is inclined to predict that it will—then Popper's bold conjecture will have been refuted.

---

[1] Let the standard deviation of $P(B)$ in the case where the B-particles are in the state $f_i(B)$ be $d_i(P(B))$. Suppose this differs from $d_{i'}(P(B))$ for some i'. Now $d(P(B))$, the standard deviation of $P(B)$ for the case at hand, where the B-particles are in a mixture of states $f_i(B)$ (with $|c_i|^2$ the probability associated with the state $f_i(B)$), is easily seen to be $\sum_i |c_i|^2 d_i(P(B))$. Hence $d(P(B))$ must be less than at least one of the $d_i(P(B))$—*viz.* the minimum one. *Q.E.D.*

2 POPPER'S PROPENSITY INTERPRETATION

In this section I shall give a general assessment of Popper's new metaphysics for QT. This may seem redundant in the light of my comments in the last section that Popper's stated views contradict QT. I shall argue however that the part of Popper's views which contradicts QT (namely the part which denied that we can ever have $S_2(B(B)) > S_1(P(B)))$ is by no means central to Popper's views. Indeed, as we have seen, Popper could well afford to agree with the CI that $S_2(P(B)) > S_1(P(B))$ on certain occasions, without prejudicing his metaphysical realism, or even introducing action-at-a-distance. I shall nevertheless reject Popper's views, but on grounds which are far more ephemeral than inconsistency with QT. I shall argue that it is the failure of Landé's programme for explaining interference effects which in the end counts against Popper's metaphysics. I shall also argue that there is a satisfactory 'realist' alternative to Popper's metaphysics, which combines the explanatory power of the CI with a sufficiently robust realism to do justice to our metaphysical concerns. These arguments, even if not convincing to Popperians, or less realist exponents of the CI, will at least illustrate that Popper has presented us with a false dichotomy: realism (Popper style) or subjectivism (CI style). Alternatives which fit into neither category are freely available and, I shall argue, suffer none of the drawbacks of either pole of Popper's dichotomy.

There are two core elements to Popper's metaphysics for QT. The first is Popper's 'realism' about the values taken by physical quantities: Popper rejects the notion (characteristic of the CI) that a physical quantity (in QT) may have an *indeterminate* value; for Popper all 'indeterminacies' are to be taken as mere uncertainties (as with Einstein), and at any time any physical quantity either has no value determinately or has a particular value determinately. The second core element to Popper's metaphysics is his 'propensity interpretation' for the probabilities of QT, which I shall now explicate following his comments on pages 348–97 of [1983] and elsewhere.

There is an everyday sense of 'propensity' (or 'tendency') in which one talks of the propensity of a certain *type* of situation to have a certain type of outcome—thus the tendency of empires to collapse, of tosses of this die to turn up six, of marriages to fail, *etc.* But propensities may also be associated with a *particular* situation having a certain type of outcome. For example we talk of the tendency which a particular empire, the British empire, say, has to collapse at a certain time—of its 'inherent instability' at that time. Moreover we can compare these tendencies: thus the Austro-Hungarian empire is said to have been less stable at the turn of the century than the British empire, *etc.*

So too we talk of the propensity of this particular die to turn up heads on the next toss, for this horse to win the next race, for that uranium atom to decay in the next five minutes, *etc.*

What is the relation between these two different sorts of propensities—the 'general propensities', of a certain *type* of set-up having a certain type of outcome, and the 'singular propensities', of a *particular* set-up having a certain type of outcome? One view is to take the singular propensities as primary, and the general propensities as secondary. For example, the general propensity of empires to collapse would then be regarded as merely a consequence of there being various singular propensities for particular empires to collapse, with the general propensity being a sort of 'average' of the singular propensities. Another view is to take general propensities as primary, with the 'singular propensities' as special cases of general propensities, and which depend moreover on our *state of knowledge*, *i.e.* they are *non-objective*. Thus the propensity of the British empire to collapse could be reformulated as a general propensity of the members of a certain class of empires to collapse, namely the class of empires which are identical in all *recognisably* relevant (recognisable to the *observer*) respects to the British empire. (*Cf* the frequency theorists' attempt to explicate single case probabilities as epistemic probabilities). Popper would not I think accept either of these last views, because, as we shall see, he is committed to the existence of genuinely objective singular propensities which are not reducible to general propensities (except perhaps in a trivial sense—see below). Indeed I shall argue that his views are close to the following, which (at least for argument's sake) I shall adopt here.

We stipulate that any 'singular propensity'—the propensity for the particular situation S to have an outcome type O, say—is really the general propensity for situations *exactly similar* to S to have the outcome O. More particularly we envisage a hypothetical set of 're-runs' of the actual world, up to and including the situation S; and it is the general propensity of possible situations of this type to have the outcome O which we then take as our singular propensity.[1] On this stipulation, although singular propensities still emerge as special cases of general propensities, they are at least *objective* (*cf* the immediately preceding view). The latter stipulation still of course leaves open the issue of whether there are in fact singular propensities, *qua* general propensities of this rather restricted and entirely hypothetical variety. For example one might still maintain that there are in fact no singular propensities, although there are general propensities: in particular that there is no singular propensity for the British empire (or any particular empire) to collapse, although there is a general propensity for empires (in general) to collapse. This last view could be justified in various ways: for example one might concede that talk of 'hypothetical re-runs' of the British empire makes sense, but that nevertheless it is not clear that there is any well-defined propensity for collapses to be exhibited over those re-runs—*i.e.* on each re-run either a collapse occurs or it does not, but there is not enough disorder among the various collapses to classify them as 'random' in the sense required for there to be a definite propensity (*qua* probability) of collapse. Or yet again it may be held that the collapse of the British empire was a deterministic process—and hence the 'propensity' for it to collapse on exactly similar re-runs is trivially 1 (*i.e.* it must recur). But a trivial propensity of this sort is no propensity at all, *i.e.* determinism precludes chance—it is not a special case of it. Yet this scepticism with respect to singular (non-trivial) propensities for collapse may be combined with the view that the collapses of empires in general are sufficiently 'random' for there to be a stable (non-trivial) long-run relative frequency—and hence propensity (*qua* probability)—with which empires collapse. (*E.g.* even though all the collapses are determined, the particular ambient circumstances which cause them vary randomly from case to case.)

---

[1] See Popper's comments in section 22 of [1983]. Implicit here is the assumption that, at least for some possible situations, sense can be made (along 'possible world lines', say) of talk of 'possible re-runs' of that situation.

Clearly there are important metaphysical issues at stake here. Where does Popper stand? Can Popper be seen as accepting the stipulation I have just made (and which, for argument's sake, I shall accept): to identify 'singular propensities' as special cases of hypothetical general—and objective—propensities? Certainly the Popper of the Post-script could have no *a priori* objections to the requisite metaphysical machinery (p. 348 [1983]); although he may of course end up rejecting it if it proves of no special use as part of a satisfactory metaphysical framework for accepted scientific theory. Moreover Popper displays no signs of ever admitting propensities which are other than 'general' in the sense just discussed—*i.e.* his propensities are always 'relative to some rule specifying the conditions for accepting a situation or experiment as a repetition of another one' (pp. 71, 72 [1982b], and p. 80 [1982b] too). And so, if he is to admit singular propensities, then it must be as a special case of general propensities. And finally, Popper is explicitly committed to the *objectivity* of propensities (section 28 of [1983]); and so must reject the view of singular propensities as merely epistemic general propensities. All of these restrictions would seem to force him into accepting something like the stipulation I have made above; although, as I also indicated above, this still leaves open the further question of whether singular propensities, so construed, exist.

On this last issue however—the existence of singular propensities—Popper is less equivocal. His explicit commitment is to statistical indeterminism for *QT* (Thesis 13 of [1982b], p. 83)—in opposition to Einstein, and 'hidden variables' theorists (like Bohm). And, in the terms I have introduced here, this is tantamount to a commitment to the existence of singular propensities—for phenomena in the domain of *QT*. That is, Popper is committed to the existence of a variation in the possible outcomes for a given experimental set-up over a set of exact hypothetical re-runs—and is committed moreover to this variation being random with a stable long-run frequency. (An instance of such an experimental set-up in the domain of *QT* would be the elapse of a certain period of time for an unstable uranium atom.) What is less clear however is whether Popper would take all general propensities as reducible to singular propensities, in the fashion discussed above. For our purposes however, it does not matter what Popper's views are on this last issue, rather the significant point is that he takes the propensities of *QT* as singular and objective, albeit special cases of general propensities. The stipulation I have introduced above is intended to provide a framework within which this can be done.

It is important not to confuse the issue of whether propensities are singular with the issue of what it is that propensities are properties of. Popper is unequivocal on this latter issue. He takes all propensities as properties of particular experimental set-ups (or, more generally, of particular 'physical situations'). Thus on page 71 [1982b] he writes: 'The probability distribution is taken to be a property of a single experiment' (also see *fn* 63 on p. 71 [1982b]). This view does not contradict Popper's other view (see above) that all properties are general, because a propensity can be a propensity for a certain type of set-up to have a certain *type* of outcome, and yet be a property of a *particular* set-up. Thus it can be a property of a *particular* toss of a *particular* coin that there is probability 1/2 for the outcome 'heads' over 'repeat' tosses of that same coin—where the 'repeats' are characterised by being the same in certain chosen respects to the particular toss in question. In short we have a *general* propensity (to come up heads on a certain type of toss) which is nevertheless a property of a *particular* experimental set-up (of die plus toss). In much the same way fragility can be a property of a *particular* piece of glass, say, and yet be a disposition for a certain type of set-up (placing the glass under a certain type of stress) to have a certain type of outcome (fracture). That is the disposition is a general disposition for instances of a certain type of experiment to have a certain type of outcome, and yet is a property of an individual (a particular piece of glass).

In advancing this last view Popper sees himself in opposition to the frequency theorist, for whom probabilities (and hence propensities) are properties of sets—indeed sequences—of similar experimental set-ups. For the frequentist the probability of a die turning up six on being tossed in a certain way is the limiting relative frequency with which six turns up over a suitable sequence of such tosses—and hence is a property of the sequence of tosses. (And 'singular propensities' are then explicated as epistemic general propensities.) Popper rightly points out that this frequentist view makes it difficult to give an account of probabilities for the outcomes of situations which occur only once, or even never (pp. 352–9 [1983]). Popper only makes a minimal modification of the frequentist view however in order to avoid this difficulty: he takes the propensity to be a property of a *particular* experimental set-up[1] (rather than a sequence of them); and to be a propensity to produce a certain outcome with a certain limiting relative frequency over a suitable hypothetical sequence of 'repeats' of the set-up:

---

[1] And only a possible one at that.

> We have to visualise the conditions as endowed with a tendency, or disposition, or propensity, to produce sequences with (particular) frequencies (my parentheses, p. 356, [1983]).

There are two points at which Popper can be criticised here. Firstly, he can be criticised for taking the existence of a limiting relative frequency (albeit over an hypothetical sequence) as necessary for the existence of a propensity. Clearly it is possible for there to be no limiting relative frequency with which 'heads' appears over an infinite and suitable 'randomised' (Bernouillian) sequence of tosses of some coin, and yet for there to be a definite probability of (and hence propensity for) 'heads' turning up. Indeed the strong law of large numbers insists on precisely such a possibility, albeit with probability zero.[1] In order to avoid this criticism I propose to modify Popper's views. Instead of taking a propensity as a propensity to produce a particular outcome with a certain relative frequency (as Popper does), I shall take the propensity simply as a propensity to produce a certain outcome—and leave it at that. Then the question of whether this propensity—quantified as a probability (see below)—is equal to a particular relative frequency is to be settled by experiment (and not *a priori* as Popper does). The best we can do, independently of experiment, is to apply the strong law of large numbers to tell us that the probability of the propensity differing from certain relative frequencies (over certain Bernouillian sequences) tends to zero as those sequences increase in length.[2]

Popper can also be criticised for taking a propensity as a property of a whole experimental set-up. For example to take the propensity of this coin to turn up 'heads' on the next toss as a property both of the method of tossing to be used next and of the coin seems a somewhat strained locution. What seems more natural to say is that the *coin* has a certain propensity to come up heads on the next toss, a propensity connected with its weight distribution, shape, *etc.*—which is then manifested (if at all) by that same coin displaying a certain statistical[3] distribution of 'heads' over a sequence of actual tosses relevantly similar to the next toss. Thus the various singular propensities—viewed as propensities of the same coin on different occasions—explain the general propensity of the coin to turn up heads, where this general propensity is manifested by the statistical distribution of heads over an actual sequence of independent tosses. It is the coin, and not the total experimental set-up, which *has* (*sic*) the propensity. And to say otherwise (as Popper does) is like saying that fragility is *not* the property of this piece of glass, but rather that it is the property of the glass together with certain experimental set-ups, *e.g.* the glass being dropped on a stone floor, struck with a hammer, *etc.* Of course I am not denying that on any particular occasion when a propensity is manifested (the glass breaks, the coin turns up heads, *etc.*), what happens is the result of factors involving the whole experimental set-up. On the contrary, although it is the glass which has the propensity to break, it is nevertheless the force with which it hit together with the internal properties of the glass which cause it to break when it does.

---

[1] Moreover Popper insists that the limiting relative frequency is equal to the probability—see continuation of the above quotation—and clearly this may fail to happen even with non-zero probability, over *finite* sequences.

[2] We can perhaps do a little better than this by introducing a detachment rule to the effect that if the propensity for an outcome of type O for a set-up of type S is large enough, then it is reasonable to accept that there will be an outcome of type O for a particular set-up of type S, provided there is no evidence to the contrary. But I cannot discuss this here.

[3] 'Chance' would be Mellor's term here [1971] (as opposed to 'propensity', which is conceived of dispositionally).

In putting forward this last view I am explicitly disagreeing with Popper:

> Propensities are then *not* properties of the particle but of the objective physical situation (*fn.* 63, p. 71 [1982b], and p. 81 too).

In defence of Popper, it must be conceded that the criteria for singling out the elements in a particular set-up to which propensities should be assigned can be somewhat equivocal (Mellor discusses this at length in [1971]). Thus Popper's decision to assign propensities to the whole set-up can be defended as a (perhaps rather extreme) ***explication*** (in Carnap's sense) of these criteria—an explication the aim of which is simply to dissolve these metaphysical ambiguities. But whereas the early Popper could have defended an explication in this way, it does not seem that the later Popper—with his concern to do justice to matters metaphysical—can. Indeed I shall later argue that Popper's own style of argument for metaphysical conclusions, can be used against his view of propensities, and in favour of the modification I have been suggesting.

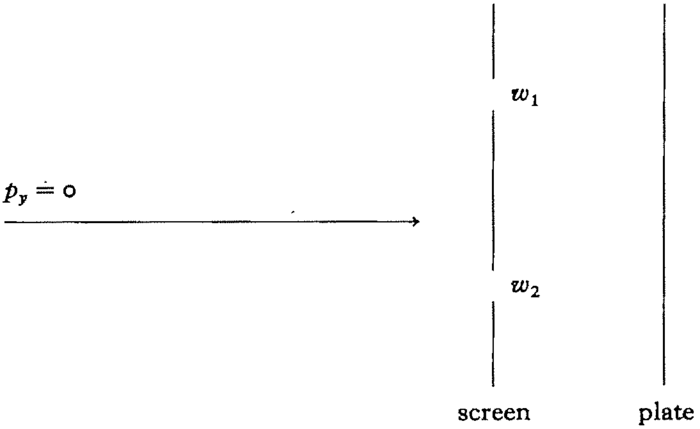
Modifying Popper's views to take into account these criticisms, we are now left with the following view of propensities. Singular propensities exist, as the objective properties of particular things—a piece of glass, a coin, or, as we shall see, the particles of *QT*. But these singular propensities can nevertheless be formulated as general propensities, *i.e.* as propensities for a certain type of outcome over exact repeats of a particular set-up. The existence of these singular propensities, in the context of *QT*, is a consequence of *QT* being indeterministic. Moreover, it seems that for Popper the existence of singular propensities presupposes the failure of determinism; so that, if coin tossing turns out ultimately to be a deterministic process, then although there may be *general* propensities for a coin to come up 'heads', there is no such singular propensity.[1] Having explicated the notion of propensity rather liberally along Popperian lines, I can now explicate Popper's 'propensity interpretation' for probabilities. Popper simply claims that certain probabilities (including those of *QT*) are nothing but quantifications of corresponding propensities. For example the singular probability (if it exists) of this coin to turn up heads on the next toss is taken to be simply a quantitative measure of the singular propensity for the next toss to have 'heads' as outcome—in the same way that a numerical 'degree of hotness' is a measure of the quality of being hot. For Popper, as we have seen, this quantitative measure of a propensity is achieved via a limiting relative frequency (over a hypothetical sequence—p. 356 [1982a]). But we need not follow him on this last point—indeed I have argued above that we should not. It is possible to assume that propensities are quantifiable just on the basis that we can *linearly order* them (*i.e.* we can say whether outcome A has a higher tendency to occur than outcome B)—as Carnap has shown in his discussion of physical magnitudes in Physics [1966]. And it is clearly possible to do this with much less machinery at our disposal than an identity between propensities and limiting relative frequencies.

---

[1] As Popper points out, even if coin tossing is a classical process, it may not be deterministic [6], Chapter II.

Against this background I shall now consider Popper's explanation of the double-slit diffraction experiment—an experiment which Popper considers central to displaying the advantages of his interpretation of QT (p. 360 [1983]). In particular I shall focus on the question of whether Popper's propensity interpretation (either in my version or his original version) makes a unique contribution to explaining quantum phenomena. Popper certainly claims that it does—and presents this claim as justification for the metaphysical assumptions involved in his propensity interpretation. So what is at issue here is, it seems, no less than the justification of Popper's whole metaphysics for QT.

Consider a stream of quantum particles (electrons) all in an eigenstate of p_y for eigenvalue 0 say (*i.e.* they all have value 0 for p_y with certainty), which move towards a screen oriented along the y-direction in which there are *two* slits occupying the y-coordinate intervals w_1 and w_2 respectively.



Some way behind the screen is a photographic plate to register the positions of the particles, after passing the screen.

Now start from the 'realist' assumption, endorsed by Popper, that each particle has a definite position and momentum at any instant. Consider the ensemble E of particles which reach the plate—and hence pass through the screen. Let $\mathrm{Prob}_E(y)$ be the probability of a particle from E registering at the point y on the plate. (For convenience we take the y-coordinate to be coarse grained, so that $\mathrm{Prob}_E(t) \neq 0$ for some y.) Formally divide E into two sub-ensembles E_1 and E_2, E_1 being the sub-ensemble of those particles which pass through w_1, and E_2 similarly. (That these two sub-ensembles exist, and exhaust E, presupposes the 'realist' assumption.) Now let 'w_1' denote the state of affairs of passing through the slit w_1, and similarly for 'w_2'. Then clearly,

$$\mathrm{Prob}_E(y) = \mathrm{Prob}_E(y \text{ and } (w_1 \text{ or } w_2))$$

(*i.e.* any member of E, which registers at y, passes through either w_1 or w_2).

$$= \mathrm{Prob}_E(y \text{ and } w_1) + \mathrm{Prob}_E(y \text{ and } w_2)$$

$$= \mathrm{Prob}_E(w_1)\mathrm{Prob}_E(y/w_1) + \mathrm{Prob}_E(w_2)\mathrm{Prob}_E(y/w_2)$$

Thus:

$$\text{(i)} \qquad \mathrm{Prob}_E(y) = P_1\mathrm{Prob}_{E_1}(y) + P_2\mathrm{Prob}_{E_2}(y)$$

(where $P_1 = \mathrm{Prob}_E(w_1)$, and P_2 similarly).

Now we perform a second experiment. We shut the slit w_2 but leave w_1 open, and then in all other respects repeat the first experiment—in particular we use an ensemble of particles which is in the same initial state as the ensemble for the first experiment. Let E' be the ensemble of particles which reach the plate in this second experiment. Then the crucial assumption is made:

$$\text{(ii)} \qquad \mathrm{Prob}_{E'}(y) = \mathrm{Prob}_{E_1}(y).$$

The justification given for this (*e.g.* see Reichenbach, pp. 24–32 [1944]), is as follows: whether or not slit w_2 is open does not affect the particles going through w_1 (unless we allow some sort of 'causal anomaly' such as action-at-a-distance). But if the path of a particle which goes through w_1 is causally independent of whether or not w_2 is open, then it should also be statistically independent of whether or not w_2 is open—which means that (ii) holds. *Q.E.D.*

We now perform a third experiment; but in this case it is w_2 which is open and w_1 closed. Let E'' be the ensemble of particles which reach the plate in this third experiment. We then assume that

$$\text{(iii)} \qquad \mathrm{Prob}_{E''}(y) = \mathrm{Prob}_{E_2}(y).$$

Substituting (ii) and (iii) into (i) we derive that

$$\text{(iv)} \qquad \mathrm{Prob}_E(y) = P_1\mathrm{Prob}_{E'}(y) + P_2\mathrm{Prob}_{E''}(y).$$

But it is easily shown experimentally that (iv) is false. *I.e.* experimentally we find that for some y, $\text{Prob}_E(y) = 0$ even though $\mathit{Prob}_{E'}(y) \neq 0$ and $\text{Prob}_{E''}(y) \neq 0$; and hence both P_1 and P_2 must be 0 (since they cannot be negative). But this means that for *all* y, $\text{Prob}_E(y) = 0$, which is experimentally falsified.[1]

The problem now is to explain the falsity of (iv)—where is the above derivation of (iv) in error? The 'orthodox' explanation is to reject the 'realist' assumption—in particular to take the particles in E as having an *indeterminate* location as they pass through the slit. Thus the very existence of E_1 and E_2, and hence (ii) and (iii), are brought into question; and the conclusion (iv) no longer follows. Moreover the indeterminacy of the location of the particles means that a particle's 'possible trajectories' can be taken as interfering with each other—as if the particle were really smeared out in wave form. And it is the equations of *QT* which then show how these interference effects produce the observed patterns which imply the failure of (iv). But Popper argues that we can solve the problem (of explaining the falsity of (iv)) in a much more modest way—namely by questioning (ii) and (iii) directly without the 'overkill' of questioning realism. In particular he claims that the usual justification for (ii) is quite spurious. It is spurious because there is no reason given why the closure of w_2 should not affect particles going through w_1. After all the closure of w_2 occurs well before the particles reach w_1; and hence there is time for the closure to cause some change in the response of the screen to the particles at w_1. Indeed the falsity of (iv) is evidence that just such an effect occurs. And Landé has provided a theory which describes how such effects can occur—and which predicts moreover that (iv) fails in just the way *QT* says it does (p. 59 [1982b]).

I have two criticisms to make of Popper's solution here. Firstly, his explanation of the falsity of (iv) depends crucially on Landé's theory. Without it his explanation clearly becomes *ad hoc* in a perjorative sense: *i.e.* the only evidence for the falsity of (ii) is the falsity of (iv) and the falsity of (iv) is explained by the falsity of (ii). *Ad hoc* explanations of this type are perhaps acceptable *faute de mieux*; but Popper needs more than this here—because he is hoping to provide a replacement for a well-entrenched alternative explanation. And the problem for Popper is that Landé's theory has by no means proved successful in other areas of *QT*. For example it has not successfully explained interference effects between differing spin states or polarisation states. Landé's theory is as yet too incomplete to constitute a serious alternative to its non-realist orthodox competitor. (On this point see Bartley's comments in *fn* 65 [1982b], p. 72.)

My second criticism of Popper's solution is advanced purely as an *ad hominem*. Popper promised us (see above) that the double slit experiment would be a crucial test for his propensity interpretation of probabilities—

---

[1] This argument is a little too swift, because the coarse-graining of the y-coordinate means that $\text{Prob}_E(y) \neq 0$ for any y for which $\text{Prob}_{E'}(y) \neq 0$ and $\text{Prob}_{E''}(y) \neq 0$. Nevertheless (iv) can be shown to be false empirically.

indeed it was for this reason that we considered it in the first place. And the way he presents his account of the failure of (iv) does indeed make it seem as if his account does essentially involve his propensity interpretation. Thus he points out that the propensities, *qua* probabilities, for the particles which pass through w_1 to end up at location y on the plate, may change if the slit at w_2 is closed; and this is so because propensities are properties of the *total* experimental arrangement. But even if we set aside the reservations I expressed above about propensities being properties of the *total* experimental arrangement, we see that Popper's remarks so far do not explain the failure of (iv). Landé's theory, or something like it, is still needed to explain how precisely the closure of the slit at w_2 affects the particles going through w_1 (as Popper himself admits). Moreover it is clear that a frequentist about probabilities could just as easily avail himself of Landé's theory as Popper does in order to explain the double-slit experiment. And the subjectivist *CI* too has no difficulties with the double-slit experiment. Thus Popper's claim that the double-slit experiment is a crucial test for his propensity interpretation is false. None of the rivals to the propensity interpretation has the slightest difficulty in explaining the falsity of (iv)! Indeed it is apparent from reading [1982c], particularly the introduction, that Popper's propensity interpretation only really comes into its own when it is used as part of Popper's devastating explanation ('dismissal' is perhaps a better term) of the 'reduction of the wave-packet' phenomenon. With this in mind, and still in pursuit of Popper's promised vindication of his propensity interpretation, I now turn to consideration of the reduction of the wave-packet.

On page 52 of [1982b], Popper writes:

> Now what I call the great quantum muddle consists in taking a distribution function, *i.e.* a statistical measure function characterising some sample space (or perhaps some population of events), and treating it as a ***physical property of the elements of the population***. . . . Now my thesis is that this muddle is widely prevalent in *QT*, as is shown by those who speak of a 'duality of particle and wave' or of 'wavicle'.
>
> For the so-called 'wave'—the Ψ-function—may be identified with . . . a *function of a probabilistic distribution function* P, where . . . $P = |\Psi|^2$. . . . On the other hand, the *element* in question has the properties of a particle. The wave shape (in configuration space) of the Ψ-function would be from this point of view a kind of accident which poses a problem to probability theory, but which has next to nothing to do with the physical properties of the particle. It is as if I were called a 'Gauss-man' . . . in order to indicate that the distribution function of my living in the South of England has a Gaussian . . . shape (in an appropriate sample space).

Then he goes on to say (in criticism of Heisenberg):

> Take a semi-transparent mirror, and assume that the probability that light will be reflected by it is 1/2. . . . Now let the experiment be carried out with a single photon. Then the probability wave packet associated with this photon will split. . . . 'After sufficient time', Heisenberg writes, 'the two parts will be separated by any distance desired. . . .' Now let us assume that we 'find' . . . that the photon . . . was reflected. 'Then the probability', he writes, 'of finding the photon in the other part of the packet *immediately becomes zero*. The experiment at the position of the reflected wave

> thus exerts a kind of action (*reduction of the wave packet*) at the distant point occupied by the transmitted wave, and one sees that this action is propagated with a velocity greater than that of light.'
>
> Now this is the great quantum muddle with a vengeance. (pp. 76, 77, [1982b], last italics mine).

From Popper's point of view one can indeed see that Heisenberg's description of the 'reduction of the wave-packet' is a muddle. Heisenberg's description depends on taking the wave-packet as a property of an individual photon. But for Popper this involves the fallacy of taking propensities (and according to Popper, the wave packet is merely a device for representing these) as the properties of an individual particle—in this case a photon. But according to Popper, the propensities—and hence the wave-packet—should be taken as propensities of a total experimental arrangement and in particular are assigned to certain repeatable aspect of the arrangement. Thus to say that the wave-packet of the photon splits after interaction with the mirror is merely a rather too colourful way[1] of saying that over a particular virtual sequence of repetitions of photon/mirror interaction, the photon is reflected on 1/2 the occasions and is transmitted on the other occasions. And to say that the wave packet is 'reduced' (*i.e.* that the transmitted part of the packet is destroyed) by the process of observing that the photon is reflected is simply a too colourful way of saying that over a virtual sequence of photons which are observed to be reflected there are no photons which are transmitted.

In short Popper says that the 'reduction of the wave packet' is not a process at all—at least not a process in the sense of a causally related sequence of states for an individual particle. In taking this view Popper dissolves, at one fell swoop, all the traditional problems associated with the 'reduction of the wave packet'—*e.g.* the existence of super-luminal velocities (see the last line in the above quotation); Wigner's proof that the reduction of the wave-packet (von Neumann style) is not consistent with *QT* under certain rather general assumptions [1952], not to mention the unsatisfactory state of affairs of having two quite different ways for systems to evolve (namely by reduction of the wave-packet *or* in conformity with the Schrödinger equation).

Unfortunately Popper's bold solution of the problems associated with the reduction of the wave-packet (by denying that there is such a process) does not work. Popper's solution ('dissolution' is a better term) of the problems has left quite unanswered—indeed has rendered *unanswerable*—how it is that the measurement process in *QT* destroys the 'interference effects' between different eigenstates of the measured quantity. To see this, consider again Heisenberg's photon—indeed consider an ensemble *E* of such photons; and measure the physical quality *P* for each such photon after it finishes interacting with the mirror. The probability of measuring *P* to have the value p say, over the ensemble E, is then compared in two quite different situations. In the first situation there is a prior measurement carried out on each photon to determine whether it was transmitted or reflected by the mirror. In the second situation, no such measurement is carried out. It is an experimental fact that the probabilities in these two situations are different; and the problem for Popper is to explain this difference. And it seems that he *cannot* explain it, because his denial that there is a reduction of the wave-packet process means that there is no objective difference between the two situations which he can use to explain the difference in probabilities!

---

[1] 'A misleading metaphor' as Popper calls it—p. 76 [1982b].

In terms of the *CI*, we can of course explain the difference in probabilities. We say that the prior measurement of the photons has destroyed the interference effects between the eigenstates ϕ_R (for the photon to be reflected) and ϕ_T (for the photon to be transmitted); and formally we represent this by saying that the initial pure state.

$$c_R\phi_R + c_T\phi_T$$

of the photons (where c_R and c_T are numerical coefficients), has been transformed by the prior measurement into a mixture of the states ϕ_R and ϕ_T, represented by the density operator

$$|c_R|^2P(\phi_R) + |c_T|^2P(\phi_T).$$

And it is the task of 'measurement theory' to give an account of just this process. But Popper can give no such account, since he denies that there is any such process. In saying this I am not of course claiming that 'measurement theory' is without its problems (although I do think many of them can be resolved). But I am pointing out that Popper is mistaken in his claim that he has dissolved the problems of the reduction of the wave-packet; and hence he can gain no support for his propensity interpretation from this source either.

So where does this leave Popper? In my view, we can rescue much of his metaphysics—including his anti-subjectivism (although not all of his realism)—by reconsidering the double-slit experiment. With the failure of Landé's theory, the *CI* is left holding the field as the provider of an explanation for the results of the double-slit experiment. (And, in my view, the *CI* is also left as the only viable explanation for the destruction of interference effects caused by measurement—although I cannot assert this here without further argument.) But the *CI* need not be given the subjectivist connotations which Popper gives it. What the *CI* demands, for explanatory purposes, is that a wave-packet is associated with each individual particle. The early Heisenberg combined this demand with a subjectivist interpretation of the wave-packet; but this extra step was, as we shall see, by no means required. And since it is this subjectivism which Popper finds so metaphysically objectionable, let us renounce it (as the later Heisenberg did), and reinterpret the wave-packet as representing some new objective properties of individual particles. This modified *CI* is still anathema to Popper:

> the wave shape . . . has next to nothing to do with the physical properties of the particle (p. 52 [1982b]).

but at least it is consistent with one of his metaphysical convictions (*viz.* objectivism), which the original *CI* was not.

But as we have seen this modified *CI* does perfectly well explain the results of the double-slit experiment provided that we retain the notion of physical quantities having indeterminate values, where it is the wave-packet of a particle which tells us which of its physical quantities have indeterminate values—according to the usual rules of the *CI*. (Moreover this modified version of the *CI* also explains the reduction of the wave-packet, although I have not shown this here.) And so this modified *CI* does in fact gain all the support from *QT* which Popper erroneously claims for his 'realist' metaphysics. Moreover, if we persist with Popper's idea that the wave-packet is simply a shorthand way of representing certain propensities (*qua* probabilities), then we see that we are committed to just the sort of propensities which I foreshadowed above—in my modified version of Popper's propensity interpretation. *I.e.* we are committed to propensities which are the properties of particular particles (just as wave-packets are), and *not* of total experimental arrangements (as Popper says). Moreover, if we follow Popper in taking *QT* to be indeterministic then these propensities become genuine singular propensities.

The metaphysics which now emerges is close to Popper's, but not identical with it. We postulate *objective singular propensities* (so far Popperian) as the properties of *particular* particles (modified Popperian), but renounce Popper's 'realist' view that physical quantities always have *determinate* values. It is the *propensities* (as represented in the wave-packet) which emerge as the fundamental properties of a particle, rather than the values taken by physical quantities. It is this objectivist hybrid of Popper's realism and the *CI* which emerges as the appropriate metaphysical framework for *QT*.[1]

HENRY KRIPS
University of Melbourne

REFERENCES

BOHM, D. and AHARNOV, Y. [1957]: *Physical Review*, **108**, p. 1070.
CARNAP, R. [1966]: *The Philosophical Foundations of Physics*. Basic Books, N.Y.
HEISENBERG, W. [1930]: *The Physical Principles of the Quantum Theory*. Dover, N.Y.

---

[1] I have to some extent simplified here, by allowing the wave-packet as the fundamental means of state-representation rather than the density operator. And I have also ignored the delicate question of whether the wave-packet is merely a short-hand representation for probabilities, rather than an independent property which is contingently related to the relevant probabilities, via *QT*.

MELLOR, D. H. [1971]: *The Matter of Chance*. Cambridge University Press.
POPPER, K. R. [1982a]: *The Open Universe*. Hutchinson.
POPPER, K. R. [1982b]: *Quantum Theory and the Schism in Physics*. Hutchinson.
POPPER, K. R. [1983]: *Realism and the Aim of Science*. Hutchinson.
REICHENBACH, H. [1944]: *Philosophical Foundations of Quantum Mechanics*. California University Press.
SCHIFF, L. [1955]: *Quantum Mechanics*. McGraw Hill.
WIGNER, E. [1952]: *Zeitschrift für Physik*, **133**, p. 101.

UNIFICATION IN SCIENCE*

1 This book consists of a collection of essays written by von Weizsäcker between 1959 and 1970. This is a sequel to a previous collection (von Weizsäcker [1957]). The present volume was published in German in 1971 and has now been translated into English. The theme of the book is an extended discussion of the relationship between the Unity of Nature, which von Weizsäcker sees as a philosophical question and the Unity of Science which von Weizsäcker extracts from the way in which the sciences, particularly physics, have developed historically up to the present time. In brief von Weizsäcker regards the Unity of Nature as a precondition for experience and uses this as a transcendental argument for the Unity of Science in a way that is closely modelled on Kantian ideas. Von Weizsäcker believes indeed that quantum theory itself can be given such a Kantian deduction as part of a more general transcendental deduction of a totally unified physics in terms of what he terms '*ur*' objects. The ultimate unified theory in physics he regards as probably provided by Heisenberg's non-linear field equation, propounded by the latter author in a series of publications in the 1960's.[1] We will comment on the details of these arguments in a moment. But first a general point. Any attempt at providing a rationalist underpinning for the latest stop-press theories in physics is a notoriously hazardous undertaking. In fact Heinsenberg's non-linear field theory is now relegated by physicists to a limbo of speculative and unfruitful theories. Substantial progress has been made in recent years in a unification programme, but it is quite different in character from Heisenberg's work.[2] This gives the book a very dated feel so far as the discussion of physics is concerned, and also throws doubt on the Kantian approach itself. However, before discussing von Weizsäcker's own views I will summarize the problem of unification as I see it myself.

2 First, what does one mean by the unity of science? There are a number[3] of senses that can be usefully distinguished.

* Review of C. F. VON WEIZSÄCKER [1980]: *The Unity of Nature*. Farrar, Straus and Giroux. £11.00, pp. x+406.

---

[1] Cf. Heisenberg [1966] for a detailed bibliography.
[2] Cf. Salam [1981] and Redhead [1983] for details of the modern approach.
[3] We modify and extend the classic discussion given in Oppenheim and Putnam [1958].

(a) *Unity of method*: The claim here is that there is a single method by means of which all the different areas of scientific enquiry are or should be conducted. Methodology here may again be subdivided as between a context of discovery and a context of justification.

(b) *Unity of language*: This is the thesis that all scientific enquiry should be couched in a canonical language.

(c) *Reductive unity*: All the different and apparently unrelated branches of science can be reduced to the language and laws of a single branch of science, where the language embraces the entities and attributes posited by the reducing branch of science, and the laws express the nomological regularities governing the behaviour of these entities.

(d) *Ontological unity*: This is the claim that behind the apparent diversity of entities and attributes that different branches of science posit as existing in the world there is on ultimate analysis only one or a limited number of actual 'things' which manifest themselves in multifarious ways.

(e) *Structural unity*: This is the thesis that science progresses by producing unified theories. Roughly speaking a unified theory is one which cannot be decomposed into two or more disjoint parts, such that the total empirical content of the original theory is simply the union of the empirical contents of the parts. A unified theory is one in which, intuitively speaking, the axioms of the theory 'work together' in a non-trivial way in producing the whole class of its empirical predictions.

(f) *Interactive unity*: This arises when two areas of science which are apparently quite independent of one another are shown to 'interact' in unexpected ways.

A few brief comments about each of these ideas. (a) is controversial. Is there *any* method? Is there a unique method? If so what is it? This is the arena of argument for the methodologists of science. (b) was propounded in various forms by the logical positivists, the canonical language being that for expressing 'observations'. It is doubtful whether anyone would still look for a unity of language in *that* sense. (c) and (d) should not necessarily be conflated. There is much controversy as to the nature of the bridging laws used in effecting reduction and which serve to link the entities and attributes of the reducing and reduced branch of science. Broadly these bridging laws can be understood either as expressing nomological coextensionalities of thing-predicates or attribute-predicates (in the terminology of Causey [1977]) or as identity statements (synthetic or analytic). It is only on the latter interpretation that reductive unity implies ontological unity. This is how the various microreductions that have occurred in the sciences are usually understood.[1] Tables and chairs, sticks and stones are 'really' collections of molecules, the molecules are 'really' collections of atoms, the atoms are 'really' collections of subatomic particles and so on. This is the road that sees elementary particle physics as the ultimate foundation for the whole of science—the bedrock for the complete reductionist hierarchy from human consciousness downwards in its most extreme manifestation. But what happens when we reach the level of elementary particles? Should we accept a number of distinct entities at an ultimate level or should we push on to *one* ultimate entity? A difficulty arises at this point. How do we distinguish different particles from the same particles with different properties. As a simple example consider the neutron and the proton. In the 1930's Heisenberg suggested thinking of the neutron and proton as *two states* of a *single* particle called a nucleon. Did this manoeuvre effect an ontological unification? The answer to this question leads one into the thorny discussion of 'natural kinds'. What criteria should be employed in distinguishing two kinds of particle from two states of one kind of particle?[1] (e) is quite a distinct question from (d). A unified theory in the sense of (e) may effect no ontological economy at all. It is a structural question that involves some difficult formal problems in attempting to give necessary and sufficient conditions for effecting this type of unification. For example it is trivial that *all* theories are unified if deductions involving unrestricted use of &-introduction are permitted. The question of whether trivialization of this sense of unification can be avoided will be returned to below. Finally (f) exploits a sense of unity not in terms of 'one-ness' but in terms of 'unitedness' or 'connectedness'. To see the difference between (d) and (f) consider two possible attitudes to conservation of energy. One attitude would be to distinguish ontologically speaking different forms of energy, mechanical, thermal, electrical *etc.* and then to allow that one sort of energy can be converted into another sort, subject to a net conservation law. This would be unification in the sense of (f). But we might also take the view that there is in fact only one form of energy, say mechanical energy, to which all the other forms are reducible on proper theoretical analysis. Or again we might want to say that there is one entity Energy with a capital *E* which manifests itself in different ways as mechanical, thermal energy *etc.* Both these latter interpretations would be unification in the sense of (d).

---

[1] See Causey [1977] for a detailed discussion of unification and microreduction. Causey takes an uncompromising stand in regarding bridging laws as synthetic identity statements.

Let us take another example of a unified theory in physics and discuss what claim is actually being made. Consider Maxwell's equation as effecting a unification of electrical and magnetic phenomena. In appropriate units, the equations read:

$$\text{curl}\,\vec{B} = \vec{J} + \frac{\partial \vec{E}}{\partial t} \qquad (1)$$

---

[1] Useful comments on this question are given by Fales [1979] who discusses the neutron-proton example and concludes (p. 364) that natural kinds 'reflect the systematically discriminable interaction governing properties which they display'. Neutrons and protons then are different natural kinds in so far as they display different electromagnetic interactions.

$$\text{curl}\,\vec{E} = -\frac{\partial \vec{B}}{\partial t} \qquad (2)$$

$$\text{div}\,\vec{E} = \rho \qquad (3)$$

$$\text{div}\,\vec{B} = 0 \qquad (4)$$

where $\vec{E}$ and $\vec{B}$ are vectors representing the electric intensity and the magnetic induction and ρ and $\vec{J}$ are the charge and current densities which act as sources of the fields. Let us denote Maxwell's theory by T. Why is it a unified theory?

First Response: Consider the theory T_B obtained by writing $\vec{E} = 0$ in T. This is a theory of purely magnetostatic phenomena. Now write T_E for the theory obtained by putting $\vec{B} = 0$, $\frac{\partial \vec{E}}{\partial t} = 0$ in T. This is a theory of purely electrostatic phenomena. If we identify a theory with the set of all the permitted solutions of its underlying equations as in Redhead [1975] then we may consider the theory $T' = T_B \bigcup T_E$ as a non-unified theory of magnetostatic and electrostatic phenomena. Now T includes of course all of T_B and of T_E but in addition these are a whole class of *new electromagnetic* phenomena which require the *full* set of Maxwell's equations for their elucidation. So we have apparently got a clear case of structural unification and of interactive unification—intuitively electric and magnetic fields mutually affect one another and some of the empirical predictions require the utilization of the full set of equations.

Second Response: Well yes, there is certainly interactive unification. But is there really structural unification. Take again the theory T_B as defined above but now in place of T_E consider a theory T'' which consists of Maxwell's equations under the condition $\vec{E} \neq 0$. Clearly $T = T_B \bigcup T''$ and this decomposition shows the theory to lack theoretical unification. This sort of trick for defeating putative theoretical unification is always available. Just divide the solution space of the theory into two disjoint parts according to any condition or criterion you care to name and you have shown the theory to be non-unified. Why then did we ever think we had achieved theoretical unification. Well, just because we regarded the four equations (1) to (4) as the underlying principles or axioms of the theory and there is no way in which all the solutions of Maxwell's equations can be derived from a simple partition of these equations. Put another way T_B and T'' in our example can each be axiomatized, but not as a *partition* of the axioms (1) to (4). This highlights the problem that a theory which is unified under one axiomatization may not be unified when considered under another axiomatization. A noteworthy attempt to deal with this problem has been made by Watkins [1983] by attempting to specify conditions under which an axiomatization could be ruled out as 'unnatural' for the purposes of assessing theoretical unification. In particular Watkins' criteria would easily eliminate our

$T_B \bigcup T''$ decomposition as involving the introduction of an inessential constraint equation $\vec{E} = 0$. But the underlying difficulty is a profound one that might lead one to conclude that the notion of structural unification is not a useful one.

But there could be analogous difficulties with some cases of interactive unification. It may be possible to eliminate interaction or coupling between quantities by expressing the basic equations in terms of newly defined quantities. An example of this would be the introduction of normal coordinates in the problem of small oscillations of coupled dynamical systems.[1] We are then faced again with an essentialist issue. Is the theory really about independent quantities (the normal coordinates) or about interacting quantities (the original coordinates subject to *coupled* equations of motion)?

Third Response. Let us return to Maxwell's equations. May these in fact not be a case of interactive unification but of ontological unification? This approach hinges on relativistic considerations. The fields $\vec{E}$ and $\vec{B}$ have no absolute significance but are transformed into linear combinations of one another when the Maxwell equations are referred to a moving frame of reference. The invariant object to which alone objective reality should be assigned is the skew-symmetric field tensor $F_{\mu\nu}$ of the familiar 4-dimensional formulation of Maxwell's equations.[2] The *two* fields $\vec{E}$ and $\vec{B}$ have been replaced by the *single* field $F_{\mu\nu}$.

In Redhead [1983] this way of understanding the unification achieved in Maxwell's equations is linked to the possibility of a *passive* interpretation of the Lorentz transformations and it is urged that analogous claims for ontological unification may not be justified where symmetry transformations are involved which do not admit of a passive but only of an active interpretation.[3]

3 Let us now return to von Weizsäcker's work. The issues raised in section 2 of this essay are ones that one might reasonably expect to see discussed in a book entitled The Unity of Nature. But here the reader would be disappointed. There are really two major strands running through von Weizsäcker's book. Firstly, there is a popular elementary exposition of the history of physics leading up to the idea of a unified elementary particle physics. There is a great deal of repetition here in different essays in the collection, and there is very little of interest to the historian or philosopher of science, particularly since the culmination of the story is now completely out of date. Secondly there is a high-falutin philosophical strand, dealing with a transcendental deduction of unified science as a precondition of experience.

---

[1] Cf. Goldstein [1980] pp. 255–8 for further details of this procedure.

[2] Cf. Møller [1972] pp. 146–8 for full details.

[3] See Redhead [1975] for further discussion of the distinction between active and passive interpretations of symmetry transformations.

Von Weizsäcker 'proves' (if you can call it a proof) that experience is grounded in the existence of time, and that time is a unity, because it does not itself change. Time involves an open future in which elementary 'alternatives' get decided. This is linked via a version of tense logic with a non-Boolean quantum logic of yes-no questions. So we get quantum theory out of the notion of time. And that is the basis for a unified physics. I do not know whether this summary does justice to von Weizsäcker's arguments because the plain fact is I find this part of von Weizsäcker's discussion almost totally opaque. Kant's *Critique of Pure Reason* is a profound work but von Weizsäcker's attempt to improve on Kant and get out quantum theory instead of Newtonian mechanics is I believe a misguided effort. The point is that what is significant in Kant is not the supposed justification of Newtonian physics but the *general* investigation of how the phenomenal world is linked to the unity of apperception. The result of this investigation may be to demonstrate the impossibility of totally chaotic experience, but leaves open the *detailed* structural characterization of a unified science. It is I believe far too much to expect that the basic principles of quantum mechanics should emerge as synthetic *a priori* principles.

The book is well produced and reasonably priced, but the style is ponderous and obscure and I cannot recommend it to readers of the Journal.

M. L. G. REDHEAD
Chelsea College, University of London

REFERENCES

CAUSEY, R. L. [1977]: *Unity of Science*. Reidel.
FALES, E. [1979]: 'Relative essentialism', *British Journal for the Philosophy of Science*, **30**, pp. 349–70.
GOLDSTEIN, H. [1980]: *Classical Mechanics*, 2nd ed. Addison-Wesley.
HEISENBERG, W. [1966]: *Introduction to the Unified Field Theory of Elementary Particles*. Interscience.
MØLLER, C. [1972]: *The Theory of Relativity*, 2nd ed. Clarendon Press.
OPPENHEIM, P. and PUTNAM, H. [1958]: 'Unity of Science as a Working Hypothesis' in H. Feigl, M. Scriven and G. Maxwell (*eds.*). *Minnesota Studies in the Philosophy of Science*, Vol. II, University of Minnesota Press, pp. 3–36.
REDHEAD, M. L. G. [1975]: 'Symmetry in Intertheory Relations'. *Synthese*, **32**, 77–112.
REDHEAD, M. L. G. [1983]: 'Quantum Field Theory for Philosophers'. To appear in *Proceedings of the 1982 Biennial Meeting of the Philosophy of Science Association*, Vol. 2.
SALAM, A. [1981]: 'Unification of the Forces' in J. H. Mulvey (*ed.*). *The Nature of Matter*. Clarendon Press, pp. 104–25.
VON WEIZSÄCKER, C. F. [1957]: *Zum Weltbild der Physik*. Hirzel.
WATKINS, J. W. N. [1983]: *Science and Scepticism*, Princeton University Press, forthcoming.

SPACE-TIME PHYSICS AND THE PHILOSOPHY OF SCIENCE*

In a lecture delivered at Göttingen in 1907, the mathematician Hermann Minkowski showed that the Principle of Relativity and the Principle of the Constancy of Light Velocity postulated by Einstein in 1905 followed from the assumption that all physical events are located in a four-dimensional Riemannian manifold, with a flat metric of signature $(+ - - -)$, in which light signals *in vacuo* describe null curves ($ds = 0$), and inertial particles describe timelike curves ($ds^2 > 0$). Such a manifold is naturally decomposed by the choice of an inertial particle into a one-dimensional time manifold represented by the particle's timelike worldline and a three-dimensional space manifold represented by any hyperplane orthogonal to the said worldline. In this fashion, the first explicit theory of 'space-time' came into being.

Einstein, who initially looked askance on the geometrical interpretation of his ideas by his former teacher Minkowski, in the end embraced it wholeheartedly, when he realised that it paved the way for a proper handling of the phenomena of gravitation, which were a major stumbling block for his Relativity theory. Indeed, the theory of gravity published by Einstein on 25 November 1915 deserves its given name of General Relativity (*GR*) only insofar as it generalises the flat Riemannian manifold of his earlier theory—now dubbed Special Relativity (*SR*)—to a Riemannian manifold of variable curvature dependent on the distribution of matter. ('Relativity', in a more familiar acceptance, is tantamount to equivalence of frames. Thus, Einstein's Principle of Relativity of 1905 implies that any pair of physical experiments carried out under homologous conditions in two labs moving inertially relatively to one another lead to homologous results. But, as Michael Friedman makes abundantly clear in his book, *GR* does not proclaim a *general* equivalence of frames, for it certainly does not entail the preposterous conclusion that one can never tell, by means of physical experiments, one arbitrarily chosen lab from another.)

In 1923 another great mathematician, the Frenchman Elie Cartan, using the latest advances in differential geometry, due to Levi-Cività [1917], Hessenberg [1917], Weyl [1918] and himself, reformulated Newton's Gravitation Theory (*NG*) as a space-time theory in which the 'arena' of physics is made to consist not indeed in a Riemannian manifold (*i.e.* one in which all curves have 'lengths'), but in a four-dimensional manifold endowed with a so-called linear connection (which selects some curves as the analogue of Euclidian straights, *i.e.* as lines that keep a steady 'direction'), dependent—like *GR*'s variable curvature—on the distribution of matter. The new advances in geometry had shown that the difference between a flat and a non-flat manifold depends altogether on the linear connection, and follows from a given Riemannian metric only in so far as the latter uniquely determines a (particular type of) connection. Since gravitational phenomena in both *GR* and *NG* are a manifestation of space-time curvature, Cartan had found in effect a common conceptual denominator for these two theories, thus exposing the myth of their incommensurability forty years before it saw the light. Finally, as has been shown, among others, by Friedman himself, in a paper written jointly with John Earman [1973], one can readily formulate Newtonian Kinematics (*NK*) as a space-time theory by suitably simplifying the postulates of Cartan's *NG*. (This may sound like hunting doves with heavy ordnance, but it is philosophically interesting, and also, as Friedman shows in his book, didactically quite useful.)

---

* Review of Michael Friedman [1983]: *Foundations of Space-Time Theories, Relativistic Physics and the Philosophy of Science*. Princeton University Press. xvi + 385 pp.
This article was written while I was a visiting fellow at the Center for Philosophy of Science of the University of Pittsburgh. I am grateful to the Center for its hospitality and its stimulating atmosphere. I also wish to thank David Malament and John Norton for their very helpful criticism of my manuscript.

Due to Reichenbach's neglect of modern geometry—pointedly noted by Weyl [1924] in his review of the former's *Axiomatik der relativistischen Raum-Zeit-Lehre*—philosophers who have picked up their understanding of Einsteinian Relativity from Reichenbach's *Philosophy of Space-Time Theory*[1]—instead of, say, from Weyl's *Space-Time-Matter*—have failed to grasp the right conceptual links between *NK*, *NG*, *SR* and *GR*, and have therefore tended to see the chasm between Einstein's thought and Newton's as much wider and deeper than it is. Reichenbach's views on space and time have been strongly criticized in recent philosophical books (Nerlich [1976], Angel [1980]), and the meagreness of their scientific foundations should be obvious to any student of a good modern standard textbook of Relativity (*e.g.* Misner, Thorne and Wheeler's *Gravitation*, Track One). But there was still a need for a work like Michael Friedman's, which provides a fairly complete and rigorous presentation of the conceptual framework of the space-time theories of mathematical physics, followed by a clear, detailed, convincing, and in some important respects original discussion of their chief implications for natural philosophy and for the very idea of a physical theory. (The latter will, one hopes, finally dispel the illusion that Relativity physics lent effective support to the philosophy of logical empiricism.) The book is designed for 'scientifically inclined philosophers and philosophically inclined scientists' (p.xi) with no more than a smattering of linear algebra and the calculus (the necessary differential geometry is explained in a very readable appendix—pp. 340—67), and should give them a sounder and more enlightened guidance to the Theory of Relativity and its philosophy than was hitherto available in works of this kind.

An introductory chapter (pp. 3—31) sets forth the general outlook and the aims of the book. These go farther than my above description may suggest, for Friedman also intends to study the interaction between the positivist and conventionalist tendencies of twentieth-century philosophy of science and the development of relativity, and to analyse some conceptual entanglements—*e.g.* between the Principle of Equivalence and the idea of general relativity—which he believes are noticeable in Einstein's work. While Friedman's writing on these historical questions (especially on the first one) is brilliant and certainly worth reading, the scholarly apparatus that he brings to bear on them is, I dare say, not sufficiently detailed to certify all his claims (at least with regard to Einstein's thought from 1905 to 1915).

---

[1] Reichenbach [1928], translated into English as *The Philosophy of Space and Time*.

The next four chapters (pp. 32—215) present the family of space-time theories and its best known members, namely, *NK*, *NG*, *SR* and *GR*. The chief common feature of space-time theories is that they conceive of physical processes and events as filling a region, or, in the limit, a hypersurface, surface, curve or point of a four-dimensional real differentiable manifold, which, following Minkowski, I shall call the theory's underlying *world*. (The prefix *world-*, as in *worldpoint*, *worldline*, *worldvelocity*, indicates appurtenance to this manifold. A space-time theory can have models built on several topologically inequivalent worlds.) Without going too far into mathematical minutiae, we can take note of some important properties of real n-dimensional manifolds:

(1) Any such manifold M is a topological space (*i.e.* a point set in which it makes sense to speak of neighbourhoods and continuity), each point of which has a neighbourhood that is topologically equivalent to an open ball in the n-dimensional Euclidian space $\mathbf{R}^n$. This means that a neighbourhood of each point P of M can be put into one-to-one bicontinuous correspondence with any open part of $\mathbf{R}^n$. Such a mapping of a neighbourhood of P in M into $\mathbf{R}^n$ is called a *chart* of M about P.

(2) The differentiable structure of M is given by a choice of such charts, subject to the twofold condition that (*a*) each point of M lies in the domain of at least one of them, and (*b*) the chosen charts are mutually compatible, so that the composition of a chart's inverse by any chart is, wherever it is defined, a smooth mapping of a part of $\mathbf{R}^n$ into $\mathbf{R}^n$. This enables one to define the concept of a smooth mapping of any differentiable manifold into another.

(3) Let P be a point of the manifold M, and let $F(P)$ denote the set of smooth real-valued functions defined on some neighbourhood of P, with the obvious algebraic structure given by the operations of addition of functions and multiplication of a function by a real number. Then, the set $F^*(P)$ of real-valued linear functions on $F(P)$ can be naturally structured as an infinite-dimensional real vector space. There are several essentially equivalent ways of selecting an n-dimensional subspace of $F^*(P)$, known as the tangent space at P, which I shall denote by T_PM. T_PM is a topological vector space isomorphic with $\mathbf{R}^n$, and can therefore readily be made into a differentiable manifold of the same dimension number as M.

The power of space-time theories rests almost entirely on the fact mentioned under (3). It follows from it that each point P of a manifold M is associated not only with the tangent space T_PM, but also with its dual, T_P^*M, *i.e.* the space of real-valued linear functions on T_PM, and with all the infinitely many vector spaces of multilinear functions with arguments drawn from T_PM and T_P^*M, and values in some specially designated vector space. The elements of these spaces are generally known as *tensors* at P, with some qualification indicative of the nature of their respective arguments and values. The collection of all tensors of a given type at every point of M is readily made into a differentiable manifold, the tensor bundle (of the said type) over M. A smooth mapping from M to this bundle, which assigns to each point of M a tensor at that point, is called a tensor field (of the type in question) on M. Another interesting differentiable manifold naturally associated with an n-dimensional manifold M is the so-called principal bundle of n-ads over M—an n-ad at a point P or M being an ordered basis of the tangent space T_PM. Since bundles over M are manifolds they also carry tensors at each point, have bundles over them and can have fields defined on them.

The space-time physicist thus has at his disposal a vast and varied array of mathematical objects from which to choose the intellectual representatives of physical realities. He can systematically compare the values of a chosen object at nearby space-time points by resorting to several differential operators, such as the Lie derivative and Cartan's exterior derivative, which are defined on arbitrary manifolds, or such as the Fermi derivative, Ricci's covariant derivative and covariant differential and Cartan's exterior covariant derivative, which are defined on manifolds endowed with some additional structure. Thus, for example, when classical electrodynamics is formulated as a space-time theory the electromagnetic field is conceived as a skew-symmetric (0, 2)-tensor field F;[1] half of Maxwell's equations are then packed into the simple formula $dF = 0$, where d stands for the exterior derivative. (The formula for the other half is no less concise, but requires further explanations.)

Tensor fields and their kin are often called 'geometric objects', but, at least from a historical point of view, some of them are more geometric than others, in so far as they can serve to specify in the underlying manifold the sort of properties and relations that have been traditionally studied by geometry. Thus, a (0, 2)-tensor field which assigns to each point of a manifold M a positive definite symmetric bilinear function on the respective tangent space (thereby making the latter into an inner product vector space) is all that is needed to define the length of curves, the size of angles, the volume of regions and the congruence of figures in M. (Such a tensor field is called a proper Riemannian metric on M. If we relax the requirement of positive definiteness we obtain the kind of improper Riemannian metric that characterises the 'geometry' of *SR* and *GR* space-times.) Likewise, the choice of a certain kind of vector-valued (0, 2)-tensor field on the principal bundle of n-ads over M, known as a connection form for M, provides the ground for distinguishing between the ordinary curves of M and its so-called geodesics or curves of constant direction (*straights* in Euclidian and Lobachevsky space), and also furnishes the additional structure required for defining the Fermi, covariant and exterior covariant derivatives on M.

---

[1] A V-valued (q, r)-tensor at P is a $(q+r)$-linear function from the Cartesian product of T_P^*M (taken q times) and T_PM (taken r times) to the vector space V. If $V = \mathbf{R}$ we simply speak of a (q, r)-tensor. Thus, a (0, 2)-tensor t at P is simply a real-valued bilinear function on $T_PM \times T_PM$. It is symmetric if, for any pair of arguments x and y, $t(x, y) = t(y, x)$; skew-symmetric, if $t(x, y) = -t(y, x)$. A (0, 2)-tensor field is skew-symmetric (resp. symmetric) if all its values are skew-symmetric (resp. symmetric).

The space-time theories mentioned above all assume that their respective worlds are structured by geometric objects in the aforesaid narrow sense. *NK*, *NG* and *SR* neatly separate them from the geometric objects in the wider sense selected to represent physical realities, which are linked to each other—but not to the former—by the respective theory's field equations. The great novelty of *GR* is that it puts all its designated geometric objects on a par—the Einstein field equations of gravity tie the (improper Riemannian) space-time metric to the stress-energy tensor field representing the distribution of matter and non-gravitational energy. J. L. Anderson [1967] expressed this difference by saying that, while the earlier space-time theories postulate both ***absolute*** and ***dynamic*** geometric objects, in *GR* all postulated objects are dynamic. Anderson uses his notion of absolute objects to give a characterisation of the symmetry group of a space-time theory, as follows: The symmetry group of a theory T is the largest group of smooth permutations with smooth inverses of the underlying world(s) of T which leave the absolute geometric objects of T invariant.[1] Anderson is thereby enabled to offer a vindication of sorts of the name 'General Relativity', based on the following vacuous truth: *GR*'s symmetry group is the largest—or 'most general'—conceivable, inasmuch as no permutation of *GR*'s underlying world(s) can fail to leave invariant every absolute object of the theory, for there are no such objects at all.

Friedman approves of Anderson's general idea and repeatedly resorts to it when comparing the several space-time theories among themselves; but he offers a characterisation of absolute objects which is more precise than the rather informal one given above and is less obscure than Anderson's. It goes as follows. We conceive a space-time theory as a class of structures—the theory's models—, each consisting of an underlying world and a list of geometric objects defined on that world and satisfying certain postulated conditions. (Note, by the way, that it is therefore possible to 'generalise' a given theory by relaxing one or more of its postulated conditions, or to specify a subtheory of it—*i.e.* a subclass of models—by adding more conditions.) The i-th geometric object of a space-time theory T is *absolute* in Friedman's sense if and only if any two models of T, $\langle M, F_1, \ldots, F_n\rangle$ and $\langle M, G_1, \ldots, G_n\rangle$, which share the same underlying world M, meet the following requirement: for every point P of M there is a diffeomorphism (a smooth one-to-one mapping with smooth inverse) f of a neighbourhood A of P onto a neighbourhood B of P such that, on the intersection of A and B, $G_i = f_* F_i$ (where f_* is the mapping induced by f on the bundle to which F_i belongs). At first sight it would seem that this mathematical characterisation cannot capture the intended, essentially physical distinction between 'absolute' and 'dynamic' objects. But on a closer look one might be more willing to concede that objects satisfying Friedman's rigid demands simply cannot serve as suitable representatives of physical fields, inextricably linked to one another in the dynamic give and take of nature. However, the characterization is not unexceptionable. On the one hand, it may be too wide, for it implies that in the sub-theory of *GR* in which matter is pressureless dust (stress-energy tensor $T^{ab} = \rho U^a U^b$) the worldvelocity of matter (U^a) is an absolute object. Friedman, who mentions this counter-example on page 59, note 9, claims that it can be avoided by preferring for the said subtheory the 'more natural' formulation in which, instead of the worldvelocity U^a *and* the density ρ, it is their product, the world-current density ρU^a, which is designated as a distinctive geometric object. But this sort of reply will no doubt smack of *ad-hoc*-ness to many readers. On the other hand, Friedman's characterisation of absolute objects may be too narrow. For suppose that inspired by a view of spatial geometry favoured by the founders of Non-Euclidian geometry, we reformulate *NK* as a space-time theory in which the underlying world M is foliated as usual by absolute time, but the spatial geometry induced on each leaf of the foliation by a degenerate Riemannian metric g on M is a geometry of constant nonpositive curvature k—the same for all times. The theory has at least one model for each admissible value of k, and, following Lobachevsky, we imagine further that it is up to experience to tell us which is the model we live in. Since g, with the stated latitude, is postulated *a priori* as a 'background' geometry, not subject to dynamical influences, one would expect it to be classified as an absolute object of the theory, and yet it does not satisfy Friedman's criterion. I grant that the liberalised *NK* I have sketched is not a plausible space-time theory; but methodological notions, such as that of an absolute object, should be liable to be extended without pain to every conceivable application and not be restricted to those which are already familiar.

---

[1] By a permutation of a set S I mean a one-to-one mapping of S onto itself. To understand how a smooth permutation of a differentiable manifold M can be said to leave invariant a given geometric object g on M, we must bear in mind that any such permutation f induces a linear isomorphism of the tangent and cotangent spaces at each point P in M onto, respectively, the tangent and cotangent spaces at $f(P)$. Let g be, for example, a (0, 2)-tensor field on M, assigning to each point P in M a bilinear function g_P on the tangent space T_PM. Let fP stand for $f(P)$, the image of P by the permutation f. Likewise, if V belongs to T_PM, the image of V by the induced isomorphism of T_PM onto $T_{fP}M$ will be denoted by fV. Then, f is said to leave g invariant if and only if, for any pair of tangent vectors V, W at any given point P of M, $g_P(V, W) = g_{fP}(fV, fW)$.

The difficulty in coming up with a satisfactory explication of Anderson's idea is not surprising, for the intended intuitive notion of absolute and dynamical objects has not much to do with the symmetries of space-time. Thus, unless the stress-energy tensor vanishes everywhere in a class of models of *GR* one would not wish to describe it as an absolute object of the respective subtheory. And yet, most known families of solutions of the Einstein field equations admit symmetries which leave the stress-energy tensor invariant. (Such symmetries are often the key to the discovery of the solutions.) I should also mention that in ordinary mathematical parlance, as opposed to Anderson's peculiar idiom, the symmetry group of an *arbitrary GR* space-time is not the extra-large group of all smooth permutations, but the extra-narrow one consisting of the identity alone. Indeed, if one insists in using the expression 'symmetry group of *GR*' as Anderson and Friedman do, one should aver that there are infinitely many such groups, *viz.* one for each family of globally diffeomorphic models of the theory.[1] I am not sure, therefore, whether Friedman has made a good investment of his time in trying to salvage Anderson's idea about absolute objects and symmetry groups, which anyway he is not using to vindicate Einstein's name for his theory of gravity (*cf.* p. 214), and which I do not think is relevant to the important points that Friedman wants to make.[2]

The two final chapters of Friedman's book deal, under the headings 'Relationalism' (pp. 216—63) and 'Conventionalism' (pp. 264—339), with some genuine and alleged philosophical consequences of Relativity theory. Let us take up the last one first. Epistemological conventionalism results from subscribing to Kant's clear-cut distinction between the sensuous and the intellectual factor in knowledge while at the same time dismissing Kant's views about the uniqueness and necessary validity of the latter. For a conventionalist such as Henri Poincaré, the conceptual apparatus by means of which the raw facts (*faits bruts*) of sense are transmuted into scientific facts (*faits scientifiques*) is to be chosen on grounds of convenience. Relativity theory has often been said to favour conventionalism in so far as it dismantled the classical system of physical geometry and chronometry that Kant had declared a permanent feature of human reason. Friedman fights this claim respectfully yet firmly. Without getting at the root of the evil, *viz.* the alleged cleft between sense and understanding, he links conventionalism to the underdetermination of general statements by particular observations. This can be of two kinds: some general statements are underdetermined because our observations are too few—thus, for example, 'there are an infinite number of distinct smooth curves through a finite set of data points' (p. 266)— but others would remain underdetermined even in the light of every possible observation. If the meaning of a statement were its method of verification, then any two sets of statements descriptive of a given field of phenomena and equally well confirmed by every possible observation in that field should be regarded as having the same meaning, so that a preference for one of them could only be a question of taste. But if we reject, as indeed we must, the verificationist criterion of meaning, underdetermination of the second kind must be resolved by appealing to rational methodological principles. As we shall soon see, in the chapter on Relationalism Friedman makes an important proposal for vindicating one such principle. In the chapter we are now considering he remarks further that underdetermination of the first kind can fare no better than the other one, for there is no established system of inductive logic.

---

[1] *Added in proof.* The two last remarks are granted, at least implicitly, by Hiskes [1984], in her newly proposed explication of Anderson's ideas. Hiskes introduces—under tortuous stipulations which I cannot review here—the notion of a *local dynamical symmetry group* of a space-time theory. She states that 'the class of local dynamical symmetry transformations' of *GR* 'may be identified as the class of all diffeomorphisms defined on [any manifold] *M*' (underlying, I presume, a model of *GR*), and contends that the relation which the local dynamical symmetry group bears to 'general relativistic theories' is 'very similar from the perspective of the theory's structure' to the relation which the Poincaré group bears to *SR* kinematics and electrodynamics. She then briefly argues against 'the claim that the Poincaré group of a special relativistic theory has a greater physical significance as a *symmetry* group than the class of local diffeomorphisms in a general relativistic theory'. I thank Adolf Grünbaum for bringing Hiskes' paper to my attention and providing me with a preprint.

[2] Another technical point which admits improvement is Friedman's distinction between the first-order and the second-order structure of a manifold endowed with a connection form. Not only is his definition of these expressions unclear, at least to me, but some statements in which he uses them can serve as premises for a valid argument with a false conclusion. Friedman writes on p. 188: 'If *M* is a general relativistic manifold then it has the same first-order structure as Minkowski space-time but not (in general) the same second-order structure.' Twelve lines later, he adds that in a manifold structured by a proper or improper Riemannian metric, 'the [covariant] derivative operator is determined by the first-order structure.' Since both *SR* (Minkowski) and *GR* space-times are structured by a Riemannian metric, the two quoted statements imply that there can be no difference at all between the covariant derivative operator on either type of space-time, inasmuch as the said operator is determined by the first-order structure which both types are said to share.

> There are not two different epistemological problems, one relating to our methodological criteria for making lower-level inductive inferences, the other relating to our criteria for choosing higher-level theories. Rather, there is only the single problem of *nondeductive* inference. We are given a finite amount of evidence, but an infinite number of incompatible hypotheses conform to this evidence. [...]. Consequently, we need additional methodological criteria, standards of simplicity, economy, and so forth, that go beyond mere conformity to evidence. These criteria can be justified only by showing that they tend to produce true hypotheses in actual scientific practice in the real world. (Hence, any justification of our actual inductive methods must itself be empirical [...] This involves us in an inevitable, but not necessarily vicious, circularity) (p. 273).

Friedman goes on to discuss the conventionalist teachings of Reichenbach and Grünbaum concerning physical geometry and the definition of simultaneity in *SR* for a given inertial frame. His treatment of the latter is particularly complete and satisfactory (pp. 165—76, 309—20). Friedman examines Reichenbach's doctrine of non-standard synchronism ($\varepsilon \neq 1/2$) with admirable forbearance, clarifying its confusions and pausing to explain such choice dainties of contemporary epistemological scholasticism as ε-Lorentz transformations. The upshot of it all, as I see it, may be briefly put thus (though Friedman would perhaps disapprove even of my minor concession to conventionalism). Under the principles of *SR*, to synchronise distant clocks at rest on an inertial frame in a manner compatible with the law of inertia and the natural order of time along the worldline of any ordinary massive material particle is tantamount to foliating *SR* spacetime into parallel spacelike hyperplanes. Hence, there is a three-parameter family of distinct synchronisations meeting the said conditions. To each member *t* of this family there is exactly one inertial frame *F* in which *t* constitutes standard synchronism. *F* and *t* are uniquely associated on strictly geometric grounds inasmuch as the leaves of the latter are Minkowski-orthogonal to the worldlines of particles at rest in the former. To synchronise clocks by *t* in a different frame *F′* amounts therefore to dating events relative to *F′* by the standard ($\varepsilon = 1/2$) time of *F*. By doing so, one arbitrarily privileges a direction—*viz.* the direction in which *F* moves uniformly relative to *F′* (= the direction of minimum ε in the chosen non-standard synchronism for *F′*)—in a space that was assumed isotropic (Euclidian) to start with. Of course, anyone can set his clocks by such a rule, if he so desires. He may have practical difficulties in keeping track of the priviledged direction, but his choice does not involve a more recondite freedom than, say, to set your watch by Greenwich Meridian Time when you live in Manitoba. Thus, the variety of ways in which events in an *SR* world can be partitioned into simultaneity classes adapted to inertial reference frames is fixed once and for all by the structure of *SR* space-time and is not enriched by Reichenbachian conventionalism. All the latter does, in effect, is to license the pairing of the inertial frames with the available—standard—synchronisms in every conceivable match (or mismatch).

Relationalism, as discussed in Friedman's book, is the doctrine, presumably initiated by Leibniz, according to which the physical reality of space and time (and space-time)—supposing they have any—must be fully accounted for by the reality of material substances (bodies?) and their mutual relations. Newton, following More and Barrow, postulated the existence of 'true, absolute or mathematical' space and time, as distinguished from the relative spaces and times that men can construct from observable bodies and physical processes. In a posthumously published manuscript, he pointed out that space and time could be neither substances nor attributes, as conceived by traditional ontology, but belonged in a category of their own. However, philosophers do not relish the multiplication of ontological categories beyond what they have learnt at school, and, having well-founded reservations against the assimilation of space and time to substances, they have tended for the most part to downgrade them to relations—or rather systems of relations—between substances. By thoroughly relativising space and time to moving bodies (frames), *SR* resolved the Newton-Leibniz dispute, as classically formulated, in favour of the latter; but only to raise again the dreaded spectre of a Newtonian absolute in the form of space-time. The space-time physicist who takes with Friedman—and with Newton and Einstein—a realist view of physics is committed to the existence of a model of his theory, *i.e.* of a four-dimensional manifold endowed with the additional structure determined on it by the geometrical objects designated by the theory. A model of a space-time theory is often referred to as a set of *events* standing in certain definite relations to each other. This might seem to give encouragement to relationalists. But the physical events a space-time is built from are twice removed from what commonly goes by that name in everyday language. In the first place, they are unextended and durationless, *viz.* pointlike events at an instant. In the second place, they may be just 'possible' events, that is to say, locations open, in a relational system, for actual physical events. To account for space-time in terms of ordinary events the relationalist must therefore deconstruct the two-tier idealisation that took us (*a*) from events extended in space and time to point-events, and (*b*) from physically occupied such points to empty ones. Friedman sees no problem in step (*a*), as he apparently believes that physical events can be realistically analysed without more ado into pointlike instantaneous ingredients. But he gives a detailed and very instructive discussion of (*b*).

The difference between the absolutist's understanding of space-time and that of a relationalist who feels no qualms about step (*a*) turns essentially on their different interpretation of the relationship between the space-time manifold and the physically real, occupied point-events. For the absolutist, space-time as such enjoys *bona fide* physical existence, and the set of occupied point-events is to be regarded simply as a proper subset of it. For the relationalist, only the occupied point-events can claim reality; their complement in space-time, the so-called 'possible events', serves only to round up the space-time structure into something less unwieldy than the set of true events and their actual relations. Thus, for the absolutist, the relation between occupied point-events is one of *inclusion*; for the relationalist, it is one of *embeddability*. In Friedman's terminology, a set A, structured by relations $R_1, \ldots, R_m$, is embeddable into a set B, structured by relations $S_1, \ldots, S_n$, if and only if (i) $m \leqslant n$, (ii) there is a one-to-one mapping q of the first m positive integers into the first n such that, for $i \leqslant m$, S_{qi} is a p-ary relation whenever R_i is p-ary, and (iii) there is a one-to-one mapping f of A into B such that, for each $i \leqslant m$ and suitable p, $R_i(x_1, \ldots, x_p)$ if and only if $S_{qi}(fx_1, \ldots, fx_p)$. The mapping f is then said to be an *embedding* of $\langle A, R_1, \ldots, R_m \rangle$ into $\langle B, S_1, \ldots, S_n \rangle$.[1] A simple example of embedding, in this sense, is the assignment of space and time coordinates to physical events. The set of the latter, as structured by some sort of topological nearness, is thereby embedded into $\mathbf{R}^4$. Note that many aspects of the standard structure of $\mathbf{R}^4$—*e.g.* the status of (0, 0, 0, 0) as the neutral element of the underlying additive group of a 4-dimensional real vector space—play no role whatsoever in the representation of events by this embedding. (It is also worth noting that most structural features of events are disregarded as well. In general, when one embeds experienced realities into a mathematical structure, the embedded structure never actually consists of the former as such, but is distilled from them by abstraction, *i.e.* by ignoring some of their features—and idealisation—*i.e.* by streamlining those that remain.)

---

[1] Note that this use of 'embedding' is not synonymous with the standard meaning of the word in differential geometry, where it is employed to designate a smooth one-to-one mapping of a differentiable manifold into another, which meets certain specific topological and algebraic conditions. To be embeddable in a space-time in this standard geometric sense, the aggregate of real events would have to constitute a differentiable manifold (diffeomorphic to a submanifold of the space-time in question)—an assumption that can never be substantiated by the kind of empirical evidence a relationalist is likely to accept.

Friedman concedes that empirical evidence concerning a collection of real events can only warrant its embedding, not its inclusion, in a model of a space-time theory. This implies that the relationalist, but not the absolutist, may draw support from such evidence. However, this does not settle the philosophical dispute between them in favour of the former. For, as Friedman notes, the inclusion of real events in a space-time structure—and hence the absolutist's understanding of the latter—can be vindicated on the strength of a rational methodological principle. It is at this point that Friedman makes his boldest and, to my mind, his most interesting philosophical move. He acknowledges—on page 242, note 14—that William Whewell anticipated it to some extent in his doctrine of the consilience of inductions; but, apart from an earlier paper by Friedman [1981], I do not find anything like it in recent epistemological literature. Let me try to summarise the gist of it. If I embed several physical situations $S_1, \dots, S_n$ into a space-time structure S, I represent the features of interest of $S_1, \dots, S_n$ by some, not necessarily the same, features of S. $S_1, \dots, S_n$ are therefore given, by their respective embeddings $f_1, \dots, f_n$, distinct, independent mathematical representations, which throw little or no light on one another. On the other hand, if the point-events discernible in $S_1, \dots, S_n$ are literally identified with points of S, the mappings $g_1, \dots, g_n$ that send the point-events of $S_1, \dots, S_n$ to those very same points regarded as elements of S are suitable restrictions of the canonical inclusion of the union of $S_1, \dots, S_n$ into S. Consequently, the representations of $S_1, \dots, S_n$ provided by $g_1, \dots, g_n$ must be consonant with each other, and we must also expect *every* structural feature of S to be relevant to the physical situations $S_1, \dots, S_n$. In so far as this requirement and these expectations are fulfilled, the theoretical structure S unifies the experienced situations $S_1, \dots, S_n$, thereby providing a way of confirming our understanding—and testing the reliability of our observations—of each of them by means of our knowledge of the others. Before illustrating this idea in detail through its application to space-time theories, Friedman gives an example drawn from a different field of physics:

> The hypotheses that collectively describe the molecular model of a gas of course receive confirmation via their explanation of the behavior of gases, but they also receive confirmation from all the other areas in which they are applied: from chemical phenomena, thermal and electric phenomena, and so on. By contrast, the purely phenomenological description of a gas—the nonliteral embeddability claim—receives confirmation from one area only: from the behavior of gases themselves.

> Hence, the theoretical description, in virtue of its far greater unifying power, is actually capable of acquiring more confirmation than is the phenomenological description.
>
> This may seem paradoxical. Since the phenomenological description of a gas is a logical consequence of the theoretical description, the degree of confirmation of the former must be at least as great as the degree of confirmation of the latter. My claim, however, is not that the phenomenological description is less well-confirmed than the theoretical description after the former is derived from the latter—this is of course impossible. Rather, the phenomenological description is less well-confirmed than it would be if it were *not* derived from the theoretical description but instead taken as primitive. The phenomenological description is better confirmed in the context of a total theory that includes the theoretical description than it is in the context of a total theory that excludes that description. This is because the theoretical description receives confirmation from indirect evidence—from chemical phenomena, thermal and electrical phenomena, and the like—which it then "transfers" to the phenomenological description. If the phenomenological description is removed from the context of higher-level theory, on the other hand, it receives confirmation only from direct evidence—from the behavior of gases themselves.
>
> In other words, a total theory rich in higher-level structure is likely to be better confirmed than a total theory staying on the phenomenological level, even though the latter theory may have precisely the same observational consequences as the former (pp. 243 ff.).

There is not much I could add by way of explanation to Friedman's own lucid words on this matter, but I still have three remarks to make concerning it. First of all, it seems to me that the identification of observable situations and processes with parts of a comprehensive major structure does not only serve to coordinate the scientific descriptions of those situations and processes and to ensure their mutual support, but works also as a guiding principle that makes it possible to achieve such description in the first place. Thus, to give a very simple example, it is practically inconceivable that Kepler could have ever thought of constructing an elliptical orbit from Tycho Brahe's data on Mars if he had not assumed as a matter of course that the planet's positions were points in an Euclidian space. Secondly, we must bear in mind that the identity of the elements of a space-time structure, and hence their 'literal identification' with real point-events, can only be determined up to an automorphism of the structure (*i.e.* up to a permutation of the underlying set that leaves all structural features invariant). Newton's denial that space and time are substances is based on the fact that they both share this feature. As he notes in the posthumous manuscript I mentioned, '***it is only by their mutual order and positions that the parts of time and space are understood to be the very same which in truth they are, and they do not possess any principle of individuation apart from that order and those positions.***'[1]

---

[1] Hall and Hall [1978], p. 100. The crucial importance of this ontological peculiarity of space-time points for Einstein's final rejection in November 1915 of his 1913/14 'hole argument' against the possibility of generally covariant gravitational field equations was pointed out to me by John Stachel in the Fall of 1980. At that time I resisted Stachel's interpretation for reasons spelled out in Torretti [1983], pp. 167 ff, but in the mean time I have been slowly won over to it. *Cf.* Stachel [1980].

Finally, I must say that I find it utterly naïve to think that the verb 'to be' as applied to a space-time structure, whose elements possess identity only up to an automorphism and can never be met in person but only through their blurred, evanescent corporeal manifestations, can mean the same as the verb 'to be' as applied to the tree in the quad or to the three ice cubes now melting in my drink. It follows from this that, if you share—as almost everybody does nowadays—the dreary belief that there is no better reality besides the reality you have to hand, you must look down upon the reality of space-time, which only gives food for thought, as being one of a lesser or weaker kind. But you ought not to boorishly rush to the conclusion that unless space-time can be formally reduced to such things as tables and beer mugs, it is no more than a fiction. *Tò ón légetai pollakhôs*—being is meant in various ways—, as Aristotle said; and within that variety there is surely enough room for the objects of sense which we enjoy and for the intelligible structures by which we seek to understand them.

ROBERTO TORRETTI
Universidad de Puerto Rico

REFERENCES

ANDERSON, J. L. [1967]: *Principles of Relativity Physics*, Academic Press.

ANGEL, R. B. [1980]: *Relativity. The Theory and its Philosophy*. Pergamon Press.

CARTAN, E. [1923 ff.]: 'Sur les variétés à connexion affine et la théorie de la relativité généralisée', *Annales de l'Ecole Normale Supérieure*, **40**, pp. 325–412, **41**, pp. 1–25, **42**, pp. 17–88.

EARMAN, J. and FRIEDMAN, M. [1973]: 'The meaning and status of Newton's Law of Inertia and the nature of gravitational forces', *Philosophy of Science*, **40**, pp. 329–59.

FRIEDMAN, M. [1981]: 'Theoretical Explanation', in *Reduction, time and reality, Studies in the philosophy of the natural sciences* (*ed.* R. Healey), Cambridge University Press, pp. 1–16.

HALL, A. R. and HALL, M. B. (*eds.*) [1978]: *Unpublished Scientific Papers of Isaac Newton*, Cambridge University Press.

HESSENBERG, G. [1917]: 'Vektorielle Begründung der Differentialgeometrie', *Mathematische Annalen*, **78**, pp. 187–217.

HISKES, A. L. D. [1984]: 'Space-time theories and symmetry groups', *Foundations of Physics*, **14**, pp. 307–32.

LEVI-CIVITA, T. [1917]: 'Nozione di parallelismo in una varietà qualunque', *Rendiconti Circolo Matematico di Palermo*, **42**, pp. 173–205.

MISNER, C. W., THORNE, K. S. and WHEELER, J. A. [1973]: *Gravitation*, W. E. Freeman.

NERLICH, G. [1976]: *The Shape of Space*, Cambridge University Press.

REICHENBACH, H. [1924]: *Axiomatik der relativistischen Raum-Zeit-Lehre*, Vieweg.

REICHENBACH, H. [1928]: *Philosophie der Raum-Zeit-Lehre*, W. de Gruyter.

STACHEL, J. [1980]: 'Einstein's search for general covariance, 1912–1915.' Talk given at [the] Ninth International Conference on General Relativity and Gravitation, Jena, DDR, 17 July 1980. Preprint.

TORRETTI, R. [1983]: *Relativity and Geometry*, Pergamon Press.

WEYL, H. [1918]: 'Reine Infinitesimalgeometrie', *Mathematische Zeitschrift*, **2**, pp. 384–411.

WEYL, H. [1924]: Review of Reichenbach, *Axiomatik der relativistischen Raum-Zeit-Lehre*, *Deutsche Literaturzeitung*, 30. Heft, Sp. 2122–8.

Reviews

GÄRDENFORS, P., HANSSON, B. and SAHLIN, N-E. (*eds*) [1983]: *Evidentiary Value: Philosophical, Judicial and Psychological Aspects of a Theory*. Lund: C. W. K. Gleerups. 120 pp. (Library of Theoria, No. 15.)

This is a collection of seven, closely-related essays on evidentiary value presented to S. Halldén on the occasion of his sixtieth birthday. They are mostly written in a charming, easy-to-read style that retains the fundamental seriousness of purpose in studying the topic. We have Basie, Daisy and Lazy, and Superintendent Eve E. Dence engaged in informative dialogues, never descending to the purely comic. Halldén should be delighted with his present.

The ideas of Halldén and his colleagues were not known to me before reading this book and since others may be equally unfamiliar with this interesting development, I will begin with a summary. (I have not seen the original papers. The description is based on the lucid material in the essays, especially Hansson's.) In the study of legal evidence there are two recognised elements: the *evidentiary theme* T, the fact to be proved; and the *evidentiary fact* E, the piece of evidence being evaluated. The present theory introduces the additional concept of an *evidentiary mechanism* C, which is a hypothesis about a causal connection between T and E. (This notation is Hansson's. Other essayists use different symbols.) The theme probability model considers $P(T|E)$, the probability of the fact given the evidence. The new suggestion is to use $P(C|E)$, the probability that the mechanism is present given the evidence. An example, to which several authors refer, due to Kahneman and Tversky is illustrative.

> Two cab companies operate in a given city, the Blue and the Green (according to the colour of the cab they run). 85% of the cabs in the city are Blue, and the remaining 15% are Green. A cab was involved in a hit-and-run accident at night. A witness later identified the cab as Green. The court tested the witness's ability to distinguish between Blue and Green cabs under night-time visibility conditions. It found that the witness was able to identify each colour correctly about 80% of the time, but confused it with the other colour about 20% of the time. What do you think are the chances that the errant cab was indeed Green, as the witness claimed?

Gärdenfors, in a fascinating dialogue, explores this example in illuminating detail. T is that the errant cab was Green. One fact E_1 is the witness's identification of the cab as Green. This is connected to T by an obvious causal mechanism C. The other fact, that only 15 per cent of cabs are Green, E_0, is not connected by a clear evidentiary mechanism to T. Kahneman and Tversky report that subjects tend to ignore E_0 and give 80 per cent as the answer. This agrees with $P(C|E_1E_0)$, whereas $P(T|E_1E_0)$ under reasonable assumptions (such as the independence of E_1 and E_0, given T and given $\bar{T}$, and the same propensity of cabs to get involved in accidents regardless of colour) is 41 per cent. That is, the argument using the evidentiary mechanism explains the subjects' behaviour better than does the probability theme approach: though Sahlin, in reporting some further experiments, suggests that even the causal view is unsatisfactory in explaining the behaviour of subjects in combining pieces of evidence. Goldsmith carefully argues that the evidentiary model better represents present judicial tradition in criminal cases than does the theme probability model. In the first essay, Ekelöf, professor of law at Uppsala, describes how he came to feel the probability theme approach was unsatisfactory, not in accord with judicial practice and suggested the consideration of a causal mechanism.

Ekelöf distinguishes between an evidentiary fact (E_1 above) and an *auxiliary fact* (E_0), the latter having no evidentiary mechanism. The distinction is nicely incorporated into the new theory. It also explains (as Hansson points out) the distinction clearly felt between two situations (a) 98 per cent of habitual criminals are in fact guilty when prosecuted for yet another crime, (b) a probability of guilt of 90 per cent based on evidence concerned solely with the defendant. A judge might convict on (b) but not on (a) even though the probability is higher in (a) than in (b).

The first difficulty the new theory faces arises when there are two evidentiary facts E_1 and E_2 with two evidentiary mechanisms C_1 and C_2: how are these two to be combined? The probability mechanism provides an answer by evaluation of $P(T|E_1E_2)$, using the probability calculus, though this is involved unless E_1 and E_2 are independent. Halldén makes certain assumptions: for example, $P(C_1|E_1) = P(C_1|E_1E_2\bar{C}_2)$, an independence condition, and $P(C_2|E_1E_2) \geqslant P(C_2|E_2)$, a concurrence axiom. From these it is easy to prove that

$$P(C_1 \cup C_2|E_1E_2) \geqslant P(C_1|E_1)+P(C_2|E_2)-P(C_1|E_1)P(C_2|E_2).$$

Hansson gives a careful discussion of these and other assumptions. (In his I, inequality, not equality, is all that is required.)

The ways in which pieces of evidence may combine is discussed by Freeling and Sahlin and leads them to consideration of Shafer's theory, also mentioned, to be dismissed as frivolous, by Levi. Shafer's theory also arises in a point mentioned by Ekelöf. Let us change the cab example a little and suppose the subject can see the true colour in 75 per cent of cases but guesses the remaining 25 per cent of the time. Allowing for the fact that he may guess correctly, the probability is at least 75 per cent that his report is right but where is the remaining 25 per cent to be distributed? This leads to the idea that probabilities need not add to one, only that $P(T)+P(\bar{T}) \leqslant 1$, as Shafer suggests.

The whole approach appears to me to be unsound. Before explaining why, let me draw your attention to a curious phenomenon. Here is a group working on an important problem, in a thoughtful way, who ignore completely an alternative approach. Well, perhaps this is permissible since one cannot forever be engaged in comparison. But when that alternative purports to prove that the group's approach is wrong, surely it is unscholarly to ignore the demonstration. Yet none of the essayists mention de Finetti's development. (Only Levi mentions the Italian, when accusing him of denying 'the intelligibility of chance', and then errs for 'chance is the parameter in my theorem' as he once said to me.) The phenomenon is not confined to the Swedish school. Other scientists ignore work that demolishes their own pet ideas. There is scope for an interesting psychological study here.

De Finetti argues, in agreement with legal opinion, that the matter at issue is the truth of the evidentiary theme T: T is uncertain. We wish to describe numerically the uncertainty felt about T in the light of the evidence, not about C. Under mild assumptions, that it would be legitimate for the Swedish school to dispute, de Finetti proves that although *one* numerical measure may be what you wish, *several* measures must combine in some rather restricted ways. These restrictions are what we call the laws of probability: convex, addition and multiplication. In other words, we must describe our uncertainty by the theme probability model. The distinction I want to drive home is that the model is not merely suggested: it is proved as a consequence of other intuitively appealing suggestions.

Now we come to an interesting point. The essayists are clearly aware of the probabilistic argument (though show no awareness of its demonstration) and find it defective. (Cases (a) and (b) above provide an example.) Ekelöf even produces his own demonstration of the probability laws. So is de Finetti wrong somewhere? I believe not. The resolution is that the essayists do not sufficiently understand probability. This is not surprising since it is a most subtle notion, not understood properly until recently, and hard to manipulate. I have carefully considered most of the counter-examples in this book to the theme probability approach and acted as Basie in Gärdenfors' dialogue: in no case does it give unsatisfactory answers. All the examples, and they are good ones, yield to the appropriate use of probability.

Consider (a) and (b) above. What is being reported in (a) is that in a large number, say 100, trials of habitual criminals, 98 resulted in the judgment of guilt. This is a frequency result and it will not be a gross distortion to treat 98 per cent as the chance of guilt. We now come to a new trial. How can we use the chance? It can only become a probability of 0.98 that this defendant is guilty if the new trial is judged exchangeable with the old ones. But can we make this judgment of exchangeability? It is a free world, your probabilities can be what you like provided they obey the rules of probability. But I would not make the judgment and hope that no judge would. Evidence should be produced for exchangeability: this would be specific to the criminal. The police could have selected any habitual criminal on a faked charge. Failure to distinguish between chance and probability lies at the heart of the confusion.

Another example concerns Hansson's detective story with Superintendent Dence. There are three suspects, Louie, Huey and Dewey. Let L be the theme that Louie is guilty. There are three pieces of evidence: one concerning footprints, f, another that Louie is wealthy, l, and one that the victim's wallet is missing, w. We require $P(L|flw)$. We are told $P(L|f) = \frac{1}{2}$. To proceed we require $P(lw|Lf)$ and more. Assuming independence, we need $P(lw|L)$, or $P(w|Ll)$ being easier to think about. This is the probability that Louie, being guilty and wealthy, would remove the wallet; the implication being that he does not need to kill for money but would only remove the wallet to mislead. Such considerations are missing from the tale which invents strange rules for the incorporation of evidence.

Another difficulty concerning two pieces of evidence and two evidentiary mechanisms is that it appears necessary to distinguish between concurring and contradictory pieces of evidence. (See the second axiom mentioned above.) It is easy to see within the probabilistic theme that the distinction is vacuous. E_1 and E_2 can separately concur in supporting the theme yet together they may contradict it. In symbols, $P(T|E_i) \geqslant P(T)$ for $i = 1, 2$, yet $P(T|E_1E_2) \leqslant P(T)$. Probability is very subtle.

Suppose the cab example is modified a little so that Green cabs get correctly identified 80 per cent of the time but Blue ones only 60 per cent. Denote this evidence E_0'. What is $P(C|E_0')$ now? The data are $P(g|G) = 0.8$, $P(b|B) = 0.6$ where the upper-case letter denotes the true colour and the lower-case the witness's report. Many people confuse $P(g|G)$ with $P(G|g)$, perhaps understandably when the errors are equal.

Another major reference omitted is Eggleston's work. He gives a penetrating discussion of the probability theme from a lawyer's viewpoint. The second edition deals with Cohen's arguments which are mentioned in the book under review.

In summary, these essays have a lot to teach us about evidentiary value but fail to be satisfactory because of the authors' lack of understanding of probability. I wish other authors would write as well as most of these essayists do.

D. V. LINDLEY

REFERENCES

COHEN, L. J. [1977]: *The Probable and the Provable*. Clarendon Press.

DE FINETTI, B. [1974/5]: *Theory of Probability*. (Two volumes) Wiley. (Translated from the Italian 1970 edition.)

EGGLESTON, R. [1983]: *Evidence, Proof and Probability*. Weidenfeld and Nicolson. (Second edition; the first was 1978.)

KAHNEMAN, D., SLOVIC, P. and TVERSKY, A. *Judgment under Uncertainty: Heuristics and Biases*. Cambridge University Press.

SHAFER, G. [1976]: *A Mathematical Theory of Evidence*. Princeton University Press.

HARMAN, P. M. [1982]: *Energy, Force, and Matter.* Cambridge University Press; and *Metaphysics and Natural Philosophy.* Harvester.

The history of physics in the eighteenth and nineteenth centuries is a fertile field of research for historians of science. Here we have two books on the subject by P. M. Harman (formerly P. M. Heimann), well known for his investigations into this area. The first is a fairly straightforward historical narrative, whereas the second aims to provide support for a particular philosophical view of science.

The essential argument of the first book is that around the middle of the nineteenth century physics came to be defined in terms of the unifying role of the emerging energy concept and the programme of mechanical explanation. This is taken to support Harman's general position that the conceptual innovations of the eighteenth and nineteenth centuries erode the image of classical physics as a monolithic Newtonian structure.

He begins by contrasting the highly mathematised physics of 1850 viewed as a comparatively unified and coherent structure, with that of the eighteenth century—speculative, non-quantitative, with mechanics on one side and heat, light, electricity and magnetism on the other. The desire to emulate the great success of rational mechanics then led to the increased mathematisation of these theories and their gradual assimilation into the framework of the mechanical view of nature. From this followed the unification characteristic of nineteenth-century physics.

In his second chapter Harman identifies four milestones along the road to unification; the Laplacean emphasis on mathematisation and the construction of a unified world view, Fourier's mathematical theory of heat, Fresnel's wave theory of light, involving a mechanical ether, and the formulation of the principle of conservation of energy which stressed the unity underlying physical phenomena and effectively completed the incorporation of heat, light *etc.* into the framework of the mechanical world view.

Chapter Three deals with the bridge between heat and mechanics provided by thermodynamics.

The relationship between the concept of the physical field and mechanical theories of the ether, and the way in which the further development of the former was so tightly bound up with the elaboration of the latter, is examined in some detail in Chapter Four.

Chapter Five is concerned with the support given to the mechanical view by a molecular theory of matter in general and the kinetic theory of gases in particular. However the latter also bore with it the seeds of destruction as doubts concerning its status undermined the coherence of the whole mechanical framework.

By 1900 these problems together with developments in field theory were a serious challenge to the mechanical view and the final chapter traces out its eventual decline.

The twin threads running through this history are mechanical explanation and the unification of phenomena. Although he discusses the various ways in which physicists used the former concept, Harman fails to comment on the meaning of the term 'unification'. It would seem reasonable to say that all attempts to construct unified theories spring from a basic drive towards ontological simplicity and economy. Yet it is easy to give examples of different kinds of unification. Redhead [1982] has distinguished the unification of optical and electromagnetic phenomena from the unification of electric and magnetic phenomena revealed through Maxwell's equations. The former involves the *identification* of two things, the optical and electromagnetic ethers. The latter, on the other hand, involves *combining* two things, the electric and magnetic vector fields, into one thing, the electromagnetic field tensor.

Some discussion of these points would have been useful given the central role played by this concept in Harman's account.

A lot of attention is paid to the relationship between mechanical representations and physical reality and also to the tension between physical and mathematical models of mechanical systems. However I feel a little more philosophical discussion of the concept of 'model' in the Introduction would have been helpful.

Harman's aim is to present the history in a form which is accessible to readers without a specialised knowledge of the maths and physics involved, and so no equations are given. This is a very laudable aim, but I did feel that the absence of certain key formulae was often more of a hindrance than a help.

These are only very minor criticisms however and the book as a whole is a very worthwhile work, deserving of a place on the bookshelves of students and teachers alike.

Unfortunately I feel that a similar accolade cannot be granted to the second book. Dr Harman has a large axe to grind and he grinds it in an exceedingly laborious and rather tiresome way. The axe in question concerns the role of metaphysics in science and the aim of the book is to provide historical support for the view that the articulation of theories rests on metaphysical as well as empirical constraints. By metaphysics Harman means '. . . the attempt to justify the conceptual rationale of a scientific theory by appeal to regulative maxims such as the law of causality or to criteria of simplicity, analogy or continuity; as well as to attempts to justify the intelligibility of a theory by an explication of the meaning of concepts of matter and force' (p. 3).

He considers in turn the physics of Newton, Leibniz, Kant, Faraday, Helmholtz and Maxwell, and tries to explicate the way they each appealed to metaphysics in both senses of the word given above.

Thus in Chapter Two he outlines the way in which metaphysical issues were generated by Newton's account of the physical basis of his theory of motion. These issues were centred around the relationship between force and matter and the justification for Newton's assumption of matter's essential passivity, as embodied in the concept of inertia.

Leibniz took the opposite line and affirmed the inherent activity of matter. Chapter Three is concerned with the emphasis Leibniz placed upon the metaphysical foundations of his dynamics and its basis in his theory of substance.

Chapter Four is concerned with Kant's attempt to expose the metaphysical assumptions underlying Newton's physics and to firmly establish it through their reappraisal.

Chapter Five deals primarily with Faraday but also takes in Priestley and Hutton. Harman's claim is that Faraday resorted to metaphysical argument in order to refute atomism and justify his theory of the spatial distribution of forces in the field. By arguing that matter filled space by its inherent forces, Faraday then tied his concept of force to a theory of material substance.

The way in which Helmholtz's energy physics was influenced by Kantian metaphysics is then examined in some detail.

The final case study analyses Maxwell's natural philosophy and in particular the disjunction he made between a physical hypothesis and a mathematical formalism. This distinction is regarded as basic to Maxwell's view of the role of mechanical analogies in science and to his understanding of matter as a mathematical entity rather than a substantial reality.

The last chapter gives, rather briefly, Harman's three-fold conclusion: (1) The theories comprising classical physics display considerable ontological diversity. (2) The problem of substance and the consequent appeal to metaphysics is basic to the conceptual structure of classical physics. (3) Metaphysical arguments play a crucial role in the articulation of scientific theories.

Harman's analyses are certainly very compelling, but one is left with the feeling that he understates the role of empirical factors a little. Metaphysical considerations undoubtedly played a key role in persuading Faraday of the reality of his lines of force and in causing him to reject action at a distance. However, it must be remembered that the development of field theory was essentially a response to difficult and deep-rooted problems in the electrical theory of the late eighteenth and early nineteenth centuries. The work of Oersted, for example, and indeed Faraday's own investigations, created a need for a new representation of the action of electric and magnetic forces (see Heilbron [1981]).

Harman is not the only one to have argued that metaphysical considerations play an important role in the construction and articulation of scientific theories (see Buchdahl [1970] and [1980]; Hesse [1978]). The question is, how big is this role? How much weight should metaphysics be given in relation to other factors? What is needed is a balanced account which considers, and gives the appropriate weight to, all the relevant factors involved.

It would also have been interesting if Harman had discussed whether or not these metaphysical speculations could be said to have furthered Newton *et al.* in their work, and whether their metaphysics could be said to have been successful. As McMullin [1978] has noted, these two questions should be kept separate. Newton's metaphysics was not particularly successful— his basic premise concerning the passivity of matter came under heavy critical fire and was eventually abandoned— yet it is difficult to imagine how he could have developed his science without metaphysical principles of some sort.

The book is subtitled 'The Problem of Substance in Classical Physics'. As a treatment of this topic it is quite simply inadequate. What the concept of substance actually meant in classical physics is not explained and no distinction is drawn between 'substance' and 'matter'. Yet such a distinction clearly existed. For Aristotle, matter was the unknowable substratum and was the constituent of substance; for Locke, these roles were reversed. However Locke's view of substance isn't even mentioned, and, in general, nothing is said about the influence which the philosophical discussions of substance had upon the physicist's conception of it. To give one more example, a knowledge of how Leibniz was committed to the notion of substance through his subject-predicate distinction is vital to an understanding of the emphasis he placed on this concept in his dynamics, but Harman chooses not to discuss this.

Finally we come to the form in which these arguments are presented. The style of writing is, quite frankly, dreadful. It is dull, fustian and repetitive. The same tired old phrases are used again and again, *e.g.* 'In seeking to establish/justify/explicate . . . '. The personal pronoun is avoided and the proper name used over and over again. Instead of designating an argument in the usual way it is given almost in full every time it is referred to. The result is some incredibly dense and turgid prose which is very difficult to read. This is a great shame because Dr Harman is a first-rate historian whose views are always worth consideration.

The common theme of these two books is the ontological and conceptual diversity of those theories usually lumped together under the label 'Newtonian physics'. However whereas the first is to be recommended without hesitation, the second manages to bury its very interesting analyses under a boring and impenetrable prose style.

STEVEN FRENCH
Chelsea College, University of London

REFERENCES

BUCHDAHL, G. [1970]: 'History of Science and criteria of choice', in *Historical and Philosophical Perspectives of Science*, (*ed.* R. H. Stuewer). Minneapolis.

BUCHDAHL, G. [1980]: 'Neo-transcendental approaches towards scientific theory appraisal', in *Science, Belief and Behaviour* (*ed.* D. H. Mellor). Cambridge.

HEILBRON, J. L. [1981]: 'The electrical field before Faraday', in *Conceptions of Ether* (*eds* G. N. Cantor and M. J. S. Hodge). Cambridge.

HESSE, M. B. [1978]: 'Action at a distance', in *The Concept of Matter in Modern Philosophy* (*ed.* E. McMullin). Notre Dame.

MCMULLIN, E. [1978]: *Newton on Matter and Activity*. Notre Dame.

REDHEAD, M. L. G. [1982]: 'Quantum field theory for philosophers' presented at the Boston colloquium for the philosophy of science, October 1982.

KATZ, J. J. [1981]: *Language and other abstract objects*. Rowman and Littlefield. Pp. ix+251.

This is one of a growing number of books in the underdeveloped area of the philosophy of linguistics. In K's view, the central question in this field is that of the object of study: 'What is a theory of a natural language [or one of language in general] a theory of?' (p. 21). He sees three answers: (i) the nominalist interpretation (they are about sound waves and scribbles, *i.e.* material objects), (ii) the conceptualist interpretation (they are about mental faculties or capacities, *i.e.* mental objects), and (iii) the Platonist interpretation (they are about sentences-in-themselves, *i.e.* mathematical objects). In a brusque turnaround from his previous writing, K defends the Platonist position. This book, then, is a defence of the view that linguistics is best seen as a science whose object of study is mathematical rather than material or mental in nature.

A systematic argument for K's position would involve establishing the following premisses: (a) abstract (mathematical) objects exist independently of material and mental objects (this is how K sees mathematical objects), (b) sentences or languages are best considered as abstract objects, and (c) sentences or languages thus conceived are the linguist's sole object of study. While K's 'principal aim is to solve the philosophical problem of whether abstract objects exist' (p. 12), (b) and (c) are assumed to be valid without argument.

(A) K's ontological position is that there are three independent kinds of reality: material, mental and mathematical. He does not specify any of the possible relations between these types except to say that mental and mathematical objects can be linked by the 'knowledge of' relation, which is simply left undefined. As well, that the atemporal, aspatial, objective, changeless universals he calls abstract objects exist 'can be demonstrated if there is at least one science in which the proper interpretation of theories takes them to be about abstract objects' (p. 12). Thus, the bulk of the book is devoted to a vain attempt at showing that the Platonist interpretation is the 'proper' one, as if this demonstrated the objective existence of something.

Essential to this argument is that the sense of 'proper' remain obscure, as it does, throughout the book. 'Proper' can thus be made to indicate a set of criteria specific to logical and mathematical theories and which, of course, no factual theory will be able to meet, *e.g.* maximal absolute simplicity and abstractness. With the cards stacked in this way, the conclusion is inevitable: the argument holds, abstract objects must exist, and linguistics is a kind of mathematics. This is what is touted as 'an approach [that] can expect to settle ontological controversies non-arbitrarily and without begging philosophical questions' (pp. 12–13)! 'Abracadabra' would surely have been more to the point.

Even if he could succeed where 2000 years of philosophy has failed, it would do K no good. His ontological 'system' is so ill-defined and inexplicit that it could hardly constitute an improvement over the vague, tacit views now accepted. Moreover, interpreting linguistic theories in terms of abstract objects unrelated to mind or matter makes all of the rest of science, from anatomy to zoology, absolutely irrelevant to linguistics. This clearly constitutes a regress for linguistics, since progress in science is showing new relations between things rather than assuming a head-in-the-sand position on old ones.

(B) K simply asserts that sentences and languages exist independently of other things. Though the burden of proof is on him, he offers no argument that sentences are not, for example, equivalence classes of utterances, *i.e.* theoretical constructs, nor 'mental' constructs of some other kind. If abstract objects are aspatial, atemporal and changeless, then invoking perception or introspection to find out about them leads to contradiction. K's solution is to make them only accessible by intuition, Kant's pure intuition, which exists, according to K, (i) because we have first-hand knowledge of its operations, and (ii) because no other faculty could supply such *a priori* knowledge (p. 193). (i) is neither necessary nor sufficient, even if it could be demonstrated: there are modes of perception (*e.g.* blindsight) that exist but of which we have no first-hand experience, just as there are modes of perception of which we can have first-hand experience but would not postulate as separate epistemological faculties. How is intuition to be distinguished from delirium, religious enlightenment or dreams? At best, (i) is a hint that there might be something to investigate. (ii) is already true by postulation, so it is circular, if not misleading, to adduce it to the argument. Finally, K seems not to realise that simply because an object can be modelled mathematically it does not follow that the object is mathematical in nature.

(C) Though K sees the central question of the philosophy of linguistics as one of determining the field's object of study, he sidesteps the problem by assuming, rather than arguing, his answer to it, since it seems to follow from the argument above that linguists study properties of sentences-in-themselves. K seems not to realise that any non-trivial explanatory theory of sentences will have to account for the mechanisms that produce and understand them. There is no reason to believe that sentences will have no noteworthy properties determined by the mechanisms that produce or understand them, though this appears to be K's view.

Perplexingly, K's position would exclude from linguistics all of Chomsky's work, since he is avowedly dealing with speakers' knowlgdge (part of psychology, for K) as well as the work of all of today's psycho-, socio- and neurolinguists along with most everyone else who traffics in non-formal aspects of languages.

K's position (and his arguments) could simply be considered amusingly inane or idiosyncratic if they weren't so dangerously antiscientific. By declaring the goals, methods and knowledge of science to be irrelevant and 'bothersome' (in spite of vague references to the 'standard methodological canons' of science), Platonism seeks to promote an antiscientific philosophy of linguistics which would put the field on a par with 'scientific' parapsychology and paraphysics. K's view is not to be rejected because of an 'empirical bias' in science, but because an alternative Platonist view of science has not been formulated and shown superior to the one in use. (Perhaps in his forthcoming 'Sciences of the Intuition' he will attempt to do so.)

Clearly, K has not solved the philosophical problem of whether abstract objects exist, nor has he shown that linguistics is a mathematical science. He has offered an alternative to the nominalist and conceptualist interpretations of linguistic theories which is neither systematically developed nor viable as an alternative to either of the other views. Moreover, it is wholly antiscientific. Any value the occasional insightful remark about Chomsky's position or the useful goal of promoting the philosophy of linguistics might otherwise have had is more than compensated for by the rest of the book. Unfortunately, some of the arguments provided may take in the unwary, and Platonist linguistics has the appeal of simplicity: there are none of the 'bothersome accoutrements' of empirical science, just intuition (which anyone can do) applied to the properties of sentences. In short, Platonism would make the philosophy of linguistics a convenience for the theoretician rather than a branch of metascience. Caveat lector!

MIKE DILLINGER
McGill University

The Mathematical Papers of Isaac Newton, Volume VIII: 1697–1722, edited by D. T. WHITESIDE with the assistance in publication of A. PRAG. Cambridge University Press, 1981. £85.00. lvi + 704 pp.

This volume is the capstone of the most fitting monument to Newton's mathematical genius. It is the last of a set of eight enormous volumes collecting all of Newton's mathematical papers that are worthy of study. (A finial for the monument, a unified index, will appear later.) The weight of this work is imposing and its matter and workings complex, but those historians of science and philosophy who are ill at ease in the presence of mathematics must not be discouraged from consulting these books.

This book's contents can be described briefly and listed. Covering Newton's last years of mathematical work, it is in the format of its predecessors. After a general introduction supplying relevant chronological and biographical information, and an analytical table of contents, there are

Newton's papers, extensively annotated and thematically collected into parts. Each part has an introduction of historical background and historical and mathematical analysis. All the papers, except for minor ones in appendices, are accompanied by literal English translations on facing pages. (To a great extent, literal translations are attainable for mathematical texts, but the translations are not word-for-word. On page 456, for example, '*philosophia naturalis*' is rendered as 'physics'; in context, this is correct.) The editorial notes, besides providing variant readings from drafts not printed and certain technical information, also on occasion recast the mathematical arguments in modern form. The papers in volume viii, part i, are: (1) solutions to the twin challenge problems posed by Johann Bernoulli I in 1697; (2) revisions of Newton's '*Tractatus de Quadratura Curvarum*', first written in 1693, for publication in 1704; (3) miscellaneous shorter mathematical writings; (4) texts, including the contemporary printed one, on the 'Method of Finite Differences' (1710); (5) an extensive revision and amplification of the '*De Quadratura*' (1712); (6) a correction, at the instigation of Bernoulli and his nephew, Niklaus, of proposition x of book ii of the *Principia*, in which Newton discusses the rôles of analysis and synthesis in the geometrical arguments of the *Principia* (1712); (7) Newton's attempted solution of a problem of Bernoulli's put forth by Leibniz as a challenge to English mathematicians (1716); (8) drafts of a short passage from the preface to an intended re-issue of the second edition of the *Principia*; and (9) some minor writings that complement the *Arithmetica Universalis*. Part ii commits two hundred and thirty pages to Newton's efforts in the dispute with Leibniz over priority in the discovery of the calculus.

These writings make this volume far from the liveliest of the set, but they suffice to prove that Newton's career in mathematics was far from over after his mental crisis of 1693. The physicist may be most fascinated by volume vi, with its drafts preliminary to the *Principia*, and the mathematician by volume i, where he can follow Newton's rapid growth to mathematical maturity and the concurrent development of the calculus. Nonetheless, volume viii does contain some work of importance and a very great deal of interest. Much of this interest lies, as always, in Dr Whiteside's notes and introductions, especially the latter, which themselves continue here in all their vigour. The intricacy and richness of these writings of Dr Whiteside's are not easily described, nor, it must be admitted, are the notes easily read, but they are intended for close study, not skimming, and constitute a major resource for the historian of seventeenth-century mathematics.

These volumes provide a model for the editing of scientific manuscripts, but they are not the product of simple editorial rules, though the goal of precision is paramount. It is to be regretted that no other large corpus of scientific manuscripts of the seventeenth century has yet been edited with such precision and with an equivalent sufficiency of annotation. Throughout this edition, one sees that the editor's most versatile tool is his own judgement. From the very dating and ordering of the texts to the choice of which versions to print in full, the edition is the product of close textual criticism and analysis. We see Newton through Dr Whiteside's eyes in the interpretation and the placing of emphases in the editorial apparatus. This work shows the scholar's talent on every page, as he gives form to what would otherwise have been the most intractable material. The reader, not without struggle or, perhaps, disagreement, is led to a fuller appreciation of Newton's thought and capabilities. In other words, this is not a mere edition; it is a work of history, too. Like an edition, it contains texts, the stuff of the history of science, and like an old-fashioned history of science, it contains many facts, of who did what, and when. But it is not just traditional positivistic history of science, either, as we have come to know it from other multi-volume works. Rather, it shows us what was thought, why, and in comparison with whatever else was being thought on similar subjects by contemporaries, and the interactions of ideas at the level of individuals, not of vague influences. It exemplifies best the discipline of the history of ideas. Dr Whiteside's approach has been described as 'old-fashioned', but it is not that, even if it is, by its very programme, internalist. It is *sui generis*. Would that it does not remain so.

These volumes are not self-contained and are not to be read without reference to Newton's published works, and, indeed, many of the papers included here are not intended to be read alone. (Occasionally, Dr Whiteside does include a contemporary printed text in an appendix.) This point hardly needs emphasising when one sees that a particular paper is a draft of a section of a well-known book, but even the account of the priority dispute is far from complete. Newton's *Commercium Epistolicum*, being a published work, is omitted. (It is now easily accessible in A. R. Hall's *Philosophers at War* (Cambridge, 1981).)

There is a danger that this book and its predecessors may be too little used because of the awe and fear with which they may inspire a reader innocent of mathematics. This would be a great pity, for they have much to offer scholars of various training. One can hardly deny that 'the Mathematical Papers' are mathematical. Much of the mathematics in them, although to some slight extent it is conceptually different from its modern forms, can be followed with only a school education in calculus and Taylor series, once the archaic vocabulary is learned. Still, only the most conceited among those whose education is thus limited would claim to understand, that is, to grasp and to appreciate, the substance, without trusting to Dr Whiteside; this reviewer—to make his qualifications clear—does not so claim. There is nevertheless other important and useful writing in these volumes, however secondary, and it is on this account that no historian of science of the period can overlook these volumes without committing a grave error of omission. First, for the scholar, the historian who places a high value on getting the facts right, there is, in the general introduction, as well as in the separate introductions and notes, a wealth of well-ordered biographical information that is often unobtainable from other secondary sources. A reader confronted with the massive reference apparatus could be justified in questioning such information only after consulting the primary documents for himself. The biographical content is greater than this, though, because of the extent of the analysis half-hidden amid the editorial apparatus. For example, quite apart from such trifles as the fate of Newton's investment in the South Sea Company (p. xviii), there is a very substantial account (pp. 48–61) of the history of Newton's error in proposition x of book ii of the *Principia*, which purports to solve the problem of determining the resistance of a medium impeding the motion of a projectile in such a way that, under a constant gravitational force, the projectile will follow a given trajectory. The story of the contest of wits and the urgent corrections to which this led is not entirely divorced from the story of the dispute over the calculus, and the issues are no more difficult to comprehend, but it lays greater claim to our attention for being about an important mathematical question (not a mere point of personal or national pride) and therefore may lead to a fuller understanding of Newton as man *and* mathematician. The weightiest reason that the non-mathematician ought not to shy away from these volumes is very simple: not all the contents are mathematical, because some are about, rather than in mathematics. This particular volume has a large number of such papers, though they are not the most engaging to be found herein.

The last quarter of the book, on the priority dispute, is the least edifying part of the entire set. It is well that it is included here for the record, and it is even a shade moving to find that at the end of the eighth volume one is brought full circle back to the invention of the calculus, of the first volume. Dr Whiteside does not present the full story of the dispute; as he says (p. 561), he just allows Newton here to speak for himself. But does anything among these papers throw light on Newton's mathematical work or on other products of his intellect? A sound case may be made that by reading the effluvia of the quarrel, we may gain greater insight into Newton's and Leibniz's understanding of such fundamental concepts as that of infinitesimals, and for this purpose, the papers Dr Whiteside includes are certainly useful. It might also be argued that we see Newton here as an historian of contemporary science, though this argument would be more a quip than a claim of substance. Apart from the human interest of the story, which has recently been told at length by Professor Hall, and from some evil-tempered discussion of key concepts and some inadvertent suggestions as to how the young Newton's papers were re-organised by the old Newton, there is little to hold us here, and historians of science have, on the whole, agreed in giving the dispute no more consideration than it merits. Though a large part of this volume, the dispute is a very small part of the *Mathematical Papers*.

We (and Newton, too) are most fortunate to have these volumes, the standard of whose scholarship is a model to all. From now on, every student of Newton's mathematics and every historian of Newtonian science will have to go to this edition, to consider its ways and be wise.

A. J. APT

STOVE, D. C. [1982]: *Popper and after. Four modern irrationalists*. Pergamon Press. Pp. viii + 116.

The author claims that Popper, Lakatos, Kuhn and Feyerabend are irrationalists in the philosophy of science. He writes: 'They doubt, or deny outright, that there can be any reason to believe any scientific theory; and *a fortiori* they doubt or deny, for example, that there has been any accumulation of knowledge in recent centuries' (p. viii). The real target, however, is Popper, and in the more philosophical chapters that constitute Part II of Stove's book, there are few references to the other three 'authors'.

The first part of the book, which bears the title 'Philosophy and the English language: how irrationalism about science is made credible', purports to be an exposé of the linguistic tricks and literary devices behind which the four authors have been able to hide their irrationalism. The reviewer should say that he finds this part of the discussion not only highly unsatisfactory, in several ways, but also distasteful. The author adopts a cat and mouse approach, of the sort that the reviewer associates with adversarial dispute rather than with truth seeking. For example, to refer to 'irrationalist' claims by Kuhn, who is not a philosopher, and by Feyerabend does not establish that Popper, who has indeed influenced them, is also so tainted; Stove is here engaged in 'guilt by association'. Moreover, no justification for the claim made on page viii that Popper is an irrationalist is given until around page 47, after the conclusion of the 'exposé'. Further, though the author denies that his primary aim is criticism of the philosophy of science of the four authors, claiming that his 'principal object is simply to answer a certain *historical* question' (p. viii), in fact he not only rejects Popperian views but endorses and appears to be concerned to articulate some aspects of an alternative ('inductivist') philosophy of science, although he does not say as much about this as might be wished.

Stove identifies the 'key premise of irrationalism' as deductivism, essentially the claim that all arguments that are not logically valid are unreasonable. A more precise statement of deductivism (Thesis *O*) runs: '*P* is a reason to believe *Q* only if the argument from *P* to *Q* is valid, or there is a validator of it which is either a necessary truth or a proposition about the observed' (p. 80). The argument leading to this identification is a variant of the argument in chapters 2–4 of Stove [1973], to which the present book contributes a twentieth-century appendix.

Notwithstanding the above identification, Stove argues that inductive scepticism, which is the claim that 'no proposition about the observed is a reason to believe a contingent proposition about the unobserved', neither follows from nor implies Thesis *O*. He does not explain how the expression 'follow from' is used, but in the standard sense according to which *Q* follows from *P* if and only if it cannot be the case both that *P* is true and *Q* is false, inductive scepticism *does* follow from *O*, by a necessary truth whose status, as necessary, is relatively uncontroversial.

Stove is highly critical of deductivism. Not only does he believe it to be false, but he also judges that it is to blame for the 'literary hall-marks' (p. 87) of the authors criticised, and for their 'levity' (p. 97), though six pages later he concedes that this term does not apply to Kuhn. Following ineffective and irrelevant abuse of Lakatos and Feyerabend, he accuses Popper of displaying 'levity' in 'always saying "daring" things that he does not mean.' For though Popper 'says . . . that there is no good reason to believe any scientific theory . . . he is not in earnest' (p. 98), and is just a 'parlour-pink' irrationalist—not a colour that the reviewer has come to associate with Popper. Other adjectives that characterise deductivists are 'mischievous' and 'frivolous', and deductivism appears to be held responsible for 'our authors' bohemian contempt for . . . success' (p. 99). In the middle of this farrago, Stove dares to include a criticism of the 'adversary posture' that he claims to find in Popperian philosophy of science. Incredible.

The proclaimed alternative to deductivism is the view of the inductive logicians, of whom Carnap and Hempel are named, but who appear to include all who deny the thesis of inductive scepticism. The writings of the inductive logicians, we are told, 'are entirely free from levity' (p. 99). Is there a moral to be drawn?

The reviewer believes Stove to be right, so far as he goes, in describing deductivism as a Popperian error, but the charges of irrationalism and inductive scepticism call for more careful consideration.

On pages 47–8, at least three strands of Popper's alleged irrationalism are mentioned:

(*i*) that there are not and cannot be such things as positive good reasons to believe a scientific theory;
(*ii*) that a scientific theory can never even be probable in relation to any observational evidence for it;

and

(*iii*) that the premise of an inductive argument is never a reason to believe its conclusion (*i.e.* inductive scepticism).

There is no doubt that Popper has frequently committed himself to (*iii*), and also, less often, to (*i*) and (*ii*). In addition to the quotations cited by Stove, the reader could consult, among other relatively recent sources, sections 13 to 15 of Popper's *Replies to my critics* in Schilpp [1974], and the first article of Popper [1972]. It should, however, be noted that none of the three strands above appears to imply, *pace* pages 3 and 47, that there has been no accumulation of knowledge during the past 400 years.

Stove does not pursue (*i*) and (*ii*), and they will not be pursued here. But neither implies that scientific theories cannot be true, a thesis clearly rejected by Popper, whose great respect for such theories as relativity, and emphasis on verisimilitude, are among the themes that make it hard to convict him of irrationalism in respect of these strands. With regard to (*iii*), *viz*. inductive scepticism, with which more than half the book is concerned, the situation is more subtle, but Stove's *ad hominem* strategy again proves unhappily disingenuous. For in a well-known article, of which only Popper's reply is cited by Stove, Lakatos, one of the gang of four criticised, had issued a plea to Popper for 'a whiff of inductivism' (Schilpp [1974] pp. 241–73 and 256*ff*.), a plea that is indeed ignored in Popper's reply. Lakatos suggested that Popper was at fault in failing 'to separate rationalism from irrationalism' (*op. cit.* p. 261), and was thus led to introduce and to defend a weak form of inductivism. Lakatos' writings are, admittedly, not wholly free from 'levity', but the reader should be warned of the distortion of much of Stove's argument caused by concealment of Lakatos' penetrating discussion of inductivism.

Lakatos agrees with Popper in rejecting both an inductivist logic of discovery and inductive logic as an *a priori* discipline. He suggests, however, that a synthetic and conjectural 'inductive principle' is needed to link tested but unrefuted theories with their verisimilitude. Although Lakatos clearly denies that a principle of induction could be a *logical* principle, he equally clearly rejects inductive scepticism. The reviewer judges him to be generally correct on this issue, and suggests that Popper has indeed failed to give an adequate account of a rational preference for one tested and unrefuted theory over another.

Another consideration which suggests that induction is reasonable, even in respect of (low-level) discovery, is the success of principles of statistical inference. Statistics works—and pays—and its empirical, though fallible, success calls for explanation. Consider such simple examples as interval estimates and tests of hypotheses. Of course, estimates such as minimum length confidence intervals and formal conclusions about the rejectability or otherwise of hypotheses follow deductively from the data and assumptions, but the relevant guides for action cannot be so gained from these calculations. (This assumes classical statistical inference; on a Bayesian approach, no non-deductive step occurs, at the cost of making the resulting decisions essentially subjective.)

The argument is thus that the success of much inductive behaviour is difficult to explain unless accepted principles of statistical inference help to provide, *pace* Popper, good but non-deductive reasons for belief and action.

No such arguments are visible at the sparse level of Stove's discussion. For him, the alternative to deductivism is inductive logic, with the word 'inductive' in scare quotes. But the real problem is with the word 'logic'. For even if inductive principles can be systematized with only a small number of synthetic parameters, they do not thereby become principles of logic, as Stove appears to suggest.

In conclusion, therefore, the reviewer agrees that Popper's somewhat extreme rejection of induction makes him vulnerable to some version of the charge of deductivism. But Stove's attack is too general to advance the argument, in contrast, for example, to the few pages by Lakatos in the article cited. Even though some of its conclusions are, in the reviewer's opinion, correct, the inadequacies and the 'guilt by association' manner of proceeding make it impossible to recommend this unfortunate book.

J. M. B. MOSS
King's College, London

REFERENCES

POPPER, K. R. [1972]: *Objective Knowledge. An evolutionary approach.* Oxford: Clarendon.
SCHILPP, P. A. (*ed.*) [1974]: *The Philosophy of Karl Popper.* La Salle, Ill.: Open Court, containing LAKATOS, I. 'Popper on demarcation and induction', and POPPER, K. R.: 'Replies to my critics'.
STOVE, D. C. [1973]: *Probability and Hume's Inductive Scepticism.* Oxford: Clarendon.

Brit. J. Phil. Sci. 35 (1984), 311 *Printed in Great Britain* 311

Announcement

BRITISH SOCIETY FOR THE HISTORY OF PHILOSOPHY

The British Society for the History of Philosophy has been formed to promote and foster all aspects of the study and teaching of the history of philosophy. It will hold annual conferences and occasional smaller meetings, and establish other channels of communication among its members and with workers in cognate fields of study.

An inaugural general conference will be held at Weetwood Hall, University of Leeds, on 12th and 13th April 1985. Persons who wish to be involved in this programme should communicate immediately with the local convener, G. MacDonald Ross, Department of Philosophy, The University, Leeds LS2 9JT.

A joint meeting with the British Society for the History of Science is planned for April 1986 with the theme "New perspectives on the interrelation between the history of science and the history of philosophy". Persons who wish to offer papers germane to this theme should communicate at an early date with J. R. Milton, Department of Humanities, Imperial College, Sherfield Building, London SW7 2AZ.

Membership for the 1984–85 session is £4 minimum (unwaged and retired members £2), payable to the British Society for the History of Philosophy. Subscriptions should be sent to, and further particulars may be obtained from, A. J. Holland or M. A. Stewart, BSHP, Department of Philosophy, University of Lancaster, Bowland College, Lancaster LA1 4YT.

GRAMMATICAL MAN

Information, Entropy, Language and Life
The Story of the Modern Revolution in Human Thought

Jeremy Campbell

'One of the most remarkable books I have read for a very long time'
— Colin Wilson

02.2504 8   320pp   £3.95
(paperback)
0 7139 15668 1   320pp
£12.95 (hard cover)

GOD AND THE NEW PHYSICS

Paul Davies

'There is a wealth of interesting ideas here'
— *The Times Literary Supplement*

02.2550 1   272pp   £3.95

MIND IN SCIENCE

An Exploration of the Significance of Ideas and of Experimental Findings in the Study of Mind and Matter, from Pre-Socratic Greece to the Present Day

Richard L. Gregory

'It deserves the highest praise' — *The Times Higher Education Supplement*

055.168 9   656pp   £7.95

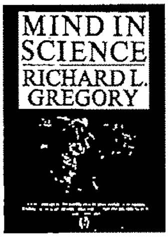
THE COSMIC CODE

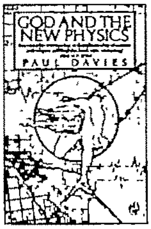
Quantum Physics as the Language of Nature

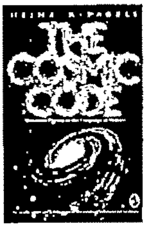
Heinz R. Pagels

A lucid and exciting guide for the layman to the world of infinitesimal particles.

02.2547 1   352pp   £4.50







For further information please write to
The Academic Marketing Department,
Penguin Books Ltd., 536 King's Road,
London SW10 0UH

penguin'84

History and Philosophy of Logic

History and Philosophy of Logic is devoted to the study of the historical development of logic and its broader philosophical concerns.

The journal is primarily concerned with general philosophical questions in logic—existential and ontological aspects, the relationship between classical and non-classical logics—including their historical development. The journal also deals with the relationships between logic and other fields of knowledge, such as mathematics, physics, philosophy of science, epistemology, linguistics, psychology and, latterly, computing.

History of Philosophy and Logic is published bi-annually in the spring and autumn. The 1984 subscription rates are £28, $71 or DM134 per volume. Back issues are also available; please apply to the publishers for details.

Taylor & Francis
London and Philadelphia

4 John Street
London WC1N 2ET, UK

NOÛS

Editor: Hector-Neri Castañeda
Associate Editor: Earl Conee

VOLUME 18, No. 2 MAY 1984

Articles: *Transcendental Arguments*/A. Brueckner; *Identity, Fuzziness and Noncontradiction*/L. Peña; *Causal Priority*/D. Hausman; *Forms and Participants in the Phaedo*/M. Matthen; *Metaquestions*/J. Driver; *Hume on Universals and General Terms*/J. Tienson.

Critical Reviews: Evan Simpson *by* James Wallace; Michael Loux *by* Nicholas Wolterstorff; Michael Resnik *by* Eike Kluge; Issac Levi *by* Henry Kyburg; Daniel Cowan *by* Swindler; Michael Devitt *by* Thomas McKay; Norman Kretzmann *by* Jóse Benardete; James Fishkin *by* Lewis Kornhauser; David Norton *by* Alasdair MacIntyre; Robert Pippin *by* Thomas Seung.

VOLUME 18 No. 3 SEPT. 1984

Articles: *Mathematical Intuition and Husserl's Phenomenology*/R. Tieszen; *A Solution to the Problem of Induction*/J. Pollock; *Hermeneutic Interpretations, Action*/G. Di Bernardo; *A Critique of Varian's Theory of Fairness*/R. Sugden.

Critical Reviews: Mary Haight *by* Robert Audi; Brian Skyrms *by* C. Hooker; Henry Teloh *by* James Forrester; Nicholas Denyer *by* Tomis Kapitan; Peter Klein *by* Ernest Sosa; B. Farrell *by* Robert Wallerstein; Eli Hirsch *by* Andrew Brennan; Thomas Donaldson *by* Richard De George.

Subscribe to:	Dept. of Philosophy	*1985 Subscriptions:*
Nous %Secretary	Indiana University	Institutions: $35.00
Submissions to:	Sycamore 126	Individuals: $18.00
Nous %Editor	Bloomington, IN 47405	Single Issues: $9 00

The Birth of History and Philosophy of Science

Kepler's *A Defence of Tycho against Ursus* with Essays on its Provenance and Significance

N. JARDINE

Introducing this first English translation of Kepler's highly original account of the history and methodology of astronomy, Dr Jardine gives an account of the work's style and strategy, its sources and the peculiar circumstances of its composition. **£32.50 net**

Theory and Measurement

HENRY E. KYBURG

Professor Kyburg proposes an original theory of the foundations of measurement, to show how quantities can be defined, why certain mathematical structures are appropriate to them and what meaning attaches to the results generated. **£22.50 net**

Cambridge Studies in Philosophy

Decisions and Revisions

Philosophical Essays on Knowledge and Value

ISAAC LEVI

This volume presents Isaac Levi's most important essays on a range of issues in epistemology, and the philosophy of science. He has developed an individual and influential view, in the tradition of Peirce and Dewey, on the questions of induction and the growth of knowledge, the foundations of probability and the theory of rational decision making. **About £25.00 net**

Forthcoming (November 1984)

Opening Pandora's Box

A Sociological Analysis of Scientists' Discourse

G. NIGEL GILBERT and **MICHAEL MULKAY**

This book provides a new approach to sociological analysis and, in particular, to the analysis of scientific culture. It moves away from previous studies which have tended to focus on scientists' actions and beliefs, to show that analysis of scientific discourse can be productive and revealing. **Hard covers £22.50 net**
Paperback £7.95 net

CAMBRIDGE UNIVERSITY PRESS

The Philosophical Quarterly

Vol.34 No 136 JULY 1984

Edited in the University of St. Andrews by an Editorial Board
Chairman of the Board of Editors: LESLIE STEVENSON
Executive Editor: GORDON GRAHAM *Reviews Editor*: ANTHONY ELLIS

Eternal Thoughts	*Peter Carruthers*
Fregean Thoughts	*Harold W. Noonan*
Frege's Platonism	*Bob Hale*
The Philosopher behind the Last Logicist	*Joan Weiner*
The Sense of a Name	*Michael Luntley*
De Re Senses	*John McDowell*
Frege, Sommers, Singular Reference	*Gregory McCulloch*
The Sense and Reference of Predicates: a Running Repair to Frege's Doctrine and a Plea for the Copula	*David Wiggins*
Frege's Metaphysical Argument	*Gregory Currie*
What does a Concept Script do?	*Cora Diamond*

Subscriptions for 1984

Annual Subscription	UK	Overseas	USA	Canada
Institutions	£22.50	£27 50	$50.00	$62.00
Individuals	£10.25	£12.50	$25.00	$30.00

Orders, remittances and enquiries should be addressed to Journals Department,
Basil Blackwell Publisher Ltd., 108 Cowley Road, Oxford OX4 1JF, England.

Philosophy and Phenomenological Research

AN INTERNATIONAL QUARTERLY FOUNDED BY MARVIN FARBER

Editors: RODERICK M. CHISHOLM and ERNEST SOSA

in cooperation with a distinguished group of American and foreign scholars

PPR publishes articles in a wide range of areas including philosophy of mind, epistemology, ethics, metaphysics, and philosophical history of philosophy. No specific methodology or philosophical orientation is required in submissions. An abstract not exceeding 250 words and two nonreturnable copies of any manuscript submitted for publication are required.

Address all communications to:
Philosophy and Phenomenological Research
BROWN UNIVERSITY, BOX 1947
PROVIDENCE, RHODE ISLAND 02912

Annual subscription rates: $15.00 (foreign, $18.00) for libraries and institutions $12.00 (foreign, $13.00) for individuals; single copies $3.75 (double issue, $7.50) and $3.00 (double issue, $6.00), respectively. There is an additional charge for postage and handling on all back issues. A circular listing the main contents of the journal and a list of available back issues will be sent upon request

EDITORIAL BOARD

R. B. Braithwaite
G. Buchdahl
L. J. Cohen
Sir John Eccles
M. B. Hesse
Sir Harold Jeffreys
C. W. Kilmister
W. C. Kneale
S. Kŏrner
D. J. O'Connor
Sir Karl Popper
H. R. Post
J. W. N. Watkins
G. J. Whitrow

EDITORIAL ADVISORY PANEL

The Editor is helped by members of the Editorial Board, members of the Committee of the British Society for the Philosophy of Science and by an Advisory Panel which includes:

Peter Clark
Jon Dorling
Arthur Fine
Allan Franklin
Ian Hacking
Colin Howson
Noretta Koertge
Hugh Mellor
David Miller
Alan Musgrave
John Worrall
Elie Zahar

THE BRITISH SOCIETY FOR THE PHILOSOPHY OF SCIENCE

Committee 1984-85

President: Dr. J. Giedymin
President-Elect: Dr. D. H. Mellor
Hon. Secretary: Dr. P. Gelpke
Hon. Treasurer: Dr. K. V. Wilkes
Members: Mr. G. M. K. Hunt, Dr. W. H. Newton-Smith,
Professor H. R. Post, Mr. A. Young

The British Journal for the Philosophy of Science is the official organ of the Society, whose purpose is described by the following statement from its Constitution and Rules:

'The purpose of the Society is to study the logic, the method and the philosophy of science as well as those of the various special sciences, including the social sciences.'

The Society holds meetings, at which papers are read and discussed, once a month during the academic year. These are held, by kind permission, in University College, London. The Committee may recognise Regional Groups of Members formed to hold meetings in other parts of the country. Any Regional Group so recognised may nominate a representative Member to the Committee.

Ordinary membership is open to all interested persons on payment of the annual subscription. Membership of the Society is international and is by no means confined to people resident in the United Kingdom. The annual subscription is £4.00 for U.K. residents and £2.00 for overseas residents. Those who wish to join should write to the Hon. Secretary, Dr. P. Gelpke, Kingston College of Further Education, Kingston Hall Road, Kingston-upon-Thames, Surrey KT1 2AQ.

Brit. J. Phil. Sci. Vol. 35, No. 3, September 1984

The British Journal for the Philosophy of Science

18 JUL 1985

Contents

Articles

201–211 RICHARD C. JENNINGS Truth, Rationality and the Sociology of Science

213–221 FRANK LAD The Calibration Question

223–243 KATHLEEN V. WILKES Is Consciousness Important?

Discussion

245–251 COLIN HOWSON Bayesianism and Support by Novel Facts

Review Articles

253–273 HENRY KRIPS Popper, Propensities, and Quantum Theory (review of K. R. POPPER: *Realism and the Aim of Science*, *The Open Universe*, and *Quantum Theory and the Schism in Physics*)

274–279 MICHAEL REDHEAD Unification in Science (review of C. F. VON WEIZSÄCKER: *The Unity of Nature*)

280–292 ROBERTO TORRETTI Space-Time Physics and the Philosophy of Science (review of MICHAEL FRIEDMAN: *Foundations of Space-Time Theories, Relativistic Physics and the Philosophy of Science*)

Reviews

293–310 D. V. Lindley on P. GÄRDENFORS, B. HANSSON, and N.-E. SAHLIN (*eds.*): *Evidentiary Value: Philosophical, Judicial and Psychological Aspects of a Theory*; Steven French on P. M. HARMAN: *Energy, Force, and Matter* and *Metaphysics and Natural Philosophy*; Mike Dillinger on J. J. KATZ: *Language and other abstract objects*; A. J. Apt on D. T. WHITESIDE (*ed.*): *The Mathematical Papers of Isaac Newton, Volume VIII: 1697–1722*; J. M. B. Moss on D. C. STOVE: *Popper and after. Four modern irrationalists.*

Announcement

311 British Society for the History of Philosophy

Printed in Great Britain at Aberdeen University Press

Volume 35 Number 4 December 1984
BJPIA 35 (No. 4) 313–419 (1984) ISSN 0007–0882

The British Journal for the Philosophy of Science

Published for the
British Society for the Philosophy of Science
by Aberdeen University Press

THE BRITISH JOURNAL FOR THE PHILOSOPHY OF SCIENCE

Editor: DR. DONALD A. GILLIES

Subscriptions

The *Journal* is published quarterly. The subscription price for mailing addresses in Britain is £14.00 net per volume post free and for mailing addresses outside Britain £17.00 (or foreign currency equivalent calculated at the day's rate); single issues £5.00 net (overseas £6.00) plus postage. Orders may be sent to any bookseller or direct to Aberdeen University Press, Farmers Hall, Aberdeen, Scotland AB9 2XT.

For members of the British Society for the Philosophy of Science—who should order direct from Aberdeen University Press—the subscription price is £9.00 (overseas £12.00). Membership of the Society is international and is by no means confined to people resident in the United Kingdom.

Microform

The current volume of the *Journal* is available in microform to subscribers to the printed version, *at a combined print/microfiche rate of* £19.50 (*overseas* £22.50).

Back Volumes

Volumes 1–16 may be obtained from Kraus Reprint Ltd., FL-9491 Nendeln, Liechtenstein or Kraus Reprint Corporation, 16 East 46 Street, New York, N.Y. 10017. Volumes 17–34 may be obtained from Aberdeen University Press, Farmers Hall, Aberdeen, Scotland AB9 2XT. Prices on application. Volumes 1–33 are also available in microform from Kraus Reprint Corporation.

Typescripts and Review Books

Typescripts, correspondence and books for review should be sent to the Editor at the Department of History and Philosophy of Science, Chelsea College, University of London, Manresa Road, London SW3 6LX.

Review books are accepted on the understanding that, although they will appear in the 'Recent Publications' list, a full review cannot be guaranteed, and that books not reviewed are not returnable.

Notes to Contributors

Three copies of all contributions must be sent. These need not, on initial submission, be in any particular style, provided they are typewritten and clearly legible. However, contributions accepted for publication may have to be wholly or partly retyped to accord with the *Journal*'s usual format—details of which may be obtained from the Editor. Contributions are considered on the assumption that they have not been published, wholly or in part, elsewhere.

Articles should be accompanied by abstracts of about 100 words.

Material is accepted for publication in the *Journal* on the condition that copyright is assigned to the British Society for the Philosophy of Science.

© British Society for the Philosophy of Science 1984

For permission to reproduce material from *The British Journal for the Philosophy of Science*, please apply to the Editor.

Brit. J. Phil. Sci. 35 (1984), 313–325 *Printed in Great Britain* 313

The Duhem Thesis

by ROGER ARIEW

1 *Duhem's Thesis is not the Duhem-Quine Thesis*
2 *Duhem's Primary Thesis is Separability, from which Falsifiability Derives*
3 *Some Consequences of Failing to Understand Duhem's Thesis*
4 *Final Comments on the Duhem Thesis*

There are numerous references in contemporary sociology of science to a thesis called the underdetermination of theories—that there are in principle an indefinite number of theories that fit the observed facts more or less adequately—attributed to Pierre Duhem and Willard Van Orman Quine (*cf.*, for example, David Bloor [1981], Karin Knorr-Cetina [1983], and Andrew Pickering [1980]). The proponents of underdetermination argue that any theory can be maintained in the face of any evidence, provided we make sufficiently radical adjustments elsewhere in our beliefs, and that from underdetermination it follows that social factors must be invoked to explain why a scientist adopts a particular theory. This general thesis (or group of interconnected theses) is complex, so that unpacking it would require some preliminaries; for now, let us just say that it is evident that these sociologists of science take what they call the underdetermination thesis (sometimes called the Duhem thesis, the D-thesis, the Duhem-Quine thesis, the Duhemian problem, and holism) as given; some even see their task at times as providing empirical evidence for it (Pickering [1980], p. 113). On the other hand, there are many references concerning the Duhem thesis by philosophers of science who are very often critical of it. For example, Noretta Koertge claims to have solved the Duhemian problem (Koertge [1978], p. 253); one can also find similar assertions by Karl Popper (Popper [1963]) and more recently by Clark Glymour and others (Glymour [1975] and [1980]).

What is curious is that Duhem himself would not have been able to recognize what is attributed to him in any of the above writings. Though curious, this is not a very unusual state of affairs, of course; often a thesis is reinterpreted according to contemporary sensibilities and attributed to its originator, even though the reinterpretation warps the original intention. But Duhem is too important a thinker to allow his thought to remain clouded by contemporary reinterpretations. Some commentators and critics have already given Duhem's thought careful exposition, and their work has demonstrated that Duhem's actual thesis is much superior to the thesis

Received 5 September 1983

normally attributed to him (*cf.* Laudan [1965]). But there is more exegetical work to be done. There are structures of Duhem's thought to be exposed that require the reading of Duhem's original text and its variations: Duhem's text spans a number of years—from 1892 to 1916, with a revised edition in 1917 (and 1981) and an English translation in 1954. The text underwent continual modifications from 1892 to 1917. In fact, it even made some difference to Duhem that he had discussed these topics as early as 1892 (1894 for the core of the Duhem thesis) since there is a question of priority of views between some of Duhem's views and the views of others—Poincaré, Milhaud, Le Roy, Wilbois, Hadamard, *etc.* (*cf.* Duhem [1954], pp. 144n, 150n, and 216n; for the publication history of Duhem's doctrine of physical theory, *cf.* Duhem [1892] to Duhem [1954]). Careful reading of Duhem's text reveals the many limitations Duhem imposed on his own thesis. I wish to exhibit these limitations and to argue for their importance to Duhem's thought. I have divided my exposition into four main sections. In section 1, I argue that Duhem's thesis is not the same as the Duhem-Quine thesis, and that Duhem's thesis cannot be faulted by the criticism that would be proper against the Duhem-Quine thesis.[1] In section 2, I argue that Duhem's primary thesis is what has been called the separability thesis, and that he regards what has been called the falsifiability thesis (which is generally taken to be his primary thesis) to be a consequence of the separability thesis. In section 3, I examine some consequences of the failure to understand Duhem's argument for alleged criticisms of the Duhem thesis. Finally, in section 4, I comment on the historical and philosophical significance of the Duhem thesis as formulated by Duhem.

1 DUHEM'S THESIS IS NOT THE DUHEM-QUINE THESIS

It is peculiar that the Duhem thesis did not make its mark on Anglo-American philosophy for a period of more than thirty years, that is, until it was introduced into Anglo-American philosophy in the 1950s through the work of W. V. O. Quine, who was relying on little more English-language sources than Phillip Frank's references to Duhem (Frank [1941]) and Armand Lowinger's book on Duhem (Lowinger [1941]). In Quine's 'Two Dogmas of Empiricism', the second dogma is reductionism, the belief that 'each meaningful statement is equivalent to some construct upon terms which refer to immediate experience' ([1953], p. 20). Quine, of course, argues that reductionism is an ill-founded dogma. He asserts that although reductionism has ceased to figure in some empiricists' thoughts, there remains a more subtle form of reductionism that each statement taken in isolation can admit of confirmation (or disconfirmation). Against this dogma Quine suggests a doctrine which he says was well argued by Pierre Duhem,

---

[1] I am extremely indebted, in this section, to Larry Laudan's excellent article on the Duhem thesis [1965]; I am also indebted to a paper he presented in April 1982 to the Philosophy Colloquium at Virginia Polytechnic Institute and State University [1982].

'that our statements about the external world face the tribunal of sense experience not individually, but only as a corporate body' ([1953], p. 41). Quine then proceeds to detail an 'empiricism without the dogmas' in which knowledge is to be likened to a field of force where 'a conflict with experience at the periphery occasions readjustments in the interior of the field' ([1953], p. 42), and 'any statement can be held true come what may, if we make drastic enough adjustments elsewhere in the system' ([1953], p. 43). Because of Quine's references to Duhem, the thesis formed from the two subtheses, that (i) since empirical statements are interconnected, they cannot be singly disconfirmed, and (ii) if we wish to hold a particular statement true we can always adjust another statement, has become known as the Duhem thesis. We should emphasise that the thesis is the conjunction of two subtheses [(i) and (ii)], and that Quine attributes only one of them to Duhem [(i)]. Duhem might have recognised the subthesis attributed to him by Quine as a step-child of his thesis, but he would *not* have approved of *either* subthesis as they are formulated above.

As previously asserted, the literature on the Duhem thesis has proliferated without regard to Duhem's formulation; some articles have attributed formulations to Duhem for which Quine might have been a more suitable attribution. For instance, Adolph Grünbaum argues that 'Duhem's thesis' is a *non-sequitur* and is actually false. The thesis that is given as Duhem's is that the '*falsifiability* of an isolated empirical hypothesis H as an *explanans* is *unavoidably inconclusive*' ([1960], p. 75; this is a paraphrase of (i)). But Grünbaum's argument against this thesis is that 'Duhem *cannot* guarantee on any logical grounds the deducibility of O [empirical findings] from an *explanans* constituted by the conjunction of H and some revised *non*-trivial version R of A [auxiliary assumptions]' ([1960], p. 75; Grünbaum also describes a case study purporting to show that Duhem's thesis is false). This argument may be an adequate argument against Quine (since it attacks subthesis (ii)), but it badly misfires against Duhem (who, it will be shown, does not hold subthesis (ii)).

It should be emphasised that Grünbaum's argument is an important argument, especially against the versions of the Duhem thesis accepted by modern relativists in sociology of science—David Bloor, Harry Collins *et al.* Lest it be thought that only sociologists of science subscribe to Quine's version of the thesis, it should be pointed out that recently Mary Hesse has defined the underdetermination thesis as the thesis that 'there are in principle always an indefinite number of theories that fit the observed facts more or less adequately' (Hesse [1980], p. viii) and has attributed the underdetermination thesis to Duhem and Quine ([1980], pp. viii and xxv). But Hesse's underdetermination thesis is either empirically false or trivially true—this is what Grünbaum said about the Duhem thesis formulated as Hesse formulates the underdetermination thesis. It is false historically that there have been an indefinite number (or a large number) of rival theories accounting for some given set of facts, and it is open empirically whether there can be an indefinite number of non-trivially different theories that can account for any given set of facts; and it is logically true that there are an indefinite number of theories that can account for any given set of facts, as long as one does not insist on the theories being non-trivially different. However, underdetermination does not need to suffer this difficulty as long as one understands underdetermination as simply the inability of empirical evidence to confirm or disconfirm conclusively a given theory—subthesis (i)—without having to add that there are an indefinite number of similar theories we can call upon to account for a given set of facts—the corollary to subthesis (ii). Of course, it is possible that by inserting the phrase 'in principle' in her formulation of the underdetermination thesis Hesse does not commit herself to there being historically (or at any time) an actual indefinite number of theories that fit the facts more or less adequately, but indicates only that she is making the logical point; though this is a possible interpretation of Hesse's thesis, one should note that Hesse also thinks that 'if we take the thesis of the underdetermination of theories seriously, relativism is a consequence that is inescapable in some form' ([1980], p. xiv). In order to arrive at such a result, the underdetermination thesis has to be more than a mere logical thesis; 'in principle' must have some empirical content. Hesse must be considering situations where in actuality there is a large number of theories that can account more or less adequately for a given set of facts; she must also be thinking that there are no criteria of rationality other than choosing the theory that accounts for the facts (or that all such criteria all together are inadequate). Given these two premises, one might be able to conclude that a relativism (most likely a sociological relativism) is needed to account for the scientist's decision for a given theory.

Having indicated that Grünbaum's argument may be an adequate argument against subthesis (ii), it should be shown that Duhem does not hold subthesis (ii). Duhem does not claim, as does the D-thesis, that when there is some conflict with experience, we can always make enough adjustments elsewhere in the system. He makes a weaker claim, also limited in other ways, but which would be equivalent to: when there is some conflict with experience what is disconfirmed is necessarily ambiguous; 'when the experiment disagrees with the physicist's prediction, he learns that at least one of the hypotheses that constitutes the set is erroneous, but the experiment does not indicate which' (Duhem [1954], p. 187).

It is clear that Duhem holds at most only a portion of the D-thesis, namely that 'if the predicted phenomenon is not produced, not only is the questioned proposition put into doubt, but also the whole theoretical scaffolding used by the physicist' (*ibid.* [1954], p. 185). There is no question here of holding any statement true come what may, merely that 'the only thing the experiment teaches us is that there is at least one error among the propositions used to predict the phenomenon and to verify that it has not been produced; but where the error lies is just what the experiment does not tell us' (*ibid.* [1954], p. 185; *cf.* also pp. 211 and 216). Some of Duhem's most telling remarks indicating that he does not hold the 'stronger' D-thesis occur in a section of *La Théorie physique* entitled 'Good Sense is the Judge of Hypotheses Which Must be Abandoned'. There Duhem refers to two possible ways of proceeding when an experiment contradicts the consequences of a theory: a timid scientist might wish to safeguard certain fundamental hypotheses and *attempt to* complicate matters by invoking various causes of error and multiplying the corrections, while a bolder scientist can *resolve to* change some of the essential suppositions supporting the entire system ([1954], pp. 216–17; emphasis mine). It is clear that for Duhem the scientist has no guarantee of success—again there is no question of upholding any hypothesis come what may: '*If they both succeed* in satisfying the requirements of the experiment, each is logically permitted to declare himself content with the work he has accomplished' ([1954], p. 217; emphasis mine). Of course, we can always reject the work of one scientist or another; 'pure logic' is not the only rule for our judgments. If 'pure logic' were the only rule for our judgments, then 'strictly speaking' we might be able to hold any statement true come what may (*cf.* Duhem [1954], p. 218, with respect to Biot's defence of the emission hypothesis against Foucault). We might think that the bold scientist has been too hasty or that the timid scientist has been unreasonably obstinate. Duhem even seems to indicate the grounds by which one can deem an alteration *ad hoc.*[1]

It has been argued that Duhem does not hold the Duhem–Quine thesis—for which more evidence will be provided later; it has also been asserted that Quine was responsible for introducing the Duhem–Quine thesis into Anglo-American philosophy. These two items together do not entail that Quine holds the Duhem–Quine thesis. That is a question open to scholarly interpretation. It is evident, at least, that Quine does not now hold the Duhem-Quine thesis as attacked by Grünbaum (*cf.* the letter to Grünbaum in Quine [1976], where Quine asserts that Grünbaum's claim 'that the Duhem–Quine thesis . . . is untenable if taken nontrivially . . . is persuasive' and that for his own part he 'would say that the thesis as [he] has used it *is* probably trivial'). It is possible that the present Quine position evolved from his earlier 'Two Dogmas of Empiricism' position; after all, Quine has now specified the structure of the field of force analogy and asserts that 'sentences higher up in theories have no empirical consequences they can call their own; they confront the tribunal of sensory evidence only in more or less inclusive aggregates. The observation sentence, situated at the sensory periphery of the body scientific, is the minimal verifiable aggregate; it has an empirical content all its own and wears it on its sleeve' (Quine [1969], p. 89); in this specification of the Quinean thesis, the sense of 'anything goes' looks like a mere logical sense. However, it is open whether this is a specification of the initial view or an evolution away from the initial view. It is entirely plausible that Quine simply overstated the initial view (*cf.* the letter to Grünbaum: 'Actually my holism is not as extreme as those two brief paragraphs at the end of "Two Dogmas of Empiricism" are bound to sound'—Quine [1976]) or that later interpreters have read Quine uncharitably.

---

[1] What Duhem indicates is that good sense is the judge of whether a hypothesis is *ad hoc.* Although a full discussion of 'good sense' would have one stray from the topic at hand, it should, at least, be indicated that by 'good sense' Duhem means an extra-logical, but not an extra-rational faculty, something like intuition, as opposed to deduction—*raison* as opposed to *raisonnement*—*cf.* Duhem [1915] where good sense is characterised as that which allows one to intuit the truth of axioms in the deductive sciences (Les Sciences de Raisonnement—Première leçon, pp. 1–22) and that which allows one to intuit fundamental principles or assumptions in experimental science (Les Sciences Expérimentales—Seconde Leçon, pp. 23–50). For a thorough discussion of good sense and Duhem's use of Pascalian ideas, *cf.* Martin [1981] (*cf.* also Martin [1982]).

2 DUHEM'S PRIMARY THESIS IS SEPARABILITY, FROM WHICH FALSIFIABILITY DERIVES

Duhem does not argue directly for what is called the Duhem thesis, or the thesis that has been taken to be his primary thesis, what Philip Quinn calls the falsifiability thesis ([1974], p. 37), even in its restricted formulation. He seems to regard it as an obvious corollary of another thesis, which, following Quinn, will be called the separability thesis (*ibid.*, [1974])[1] that the physicist can never submit an isolated hypothesis to experimental test: 'To seek to separate each of the hypotheses of theoretical physics from the other assumptions upon which this science rests, in order to subject it in isolation to observational tests, is to pursue a chimera' (Duhem [1954], pp. 199–200 and elsewhere).

Quinn has argued that the separability thesis is entailed by the falsifiability thesis, but not vice-versa (although he conjectured that Duhem might have thought the two theses entail each other) (Quinn [1974], p. 37). But he retracted this opinion following a discussion with Nancy Tuana (Tuana [1978], pp. 456–62 and Quinn [1978], pp. 463–5). Quinn argued that falsifiability does not follow from separability alone, but from separability together with the thesis that no theoretical hypothesis can be conclusively verified (if one can conclusively verify a theoretical hypothesis, then there is a falsifiable hypothesis—namely A, the hypothesis falsified by the negative experimental result affecting the conjunction of the verified hypothesis and A). Tuana responded by indicating that if any theoretical hypothesis can be conclusively verified, then the separability thesis itself is false—an excellent criticism since Quinn was attempting to give an example where separability is true and falsifiability is false by itself. Quinn now believes that the only conclusion one can draw from *La Théorie physique* is that the two theses hang together at the core of Duhem's system (Quinn, [1978]). This is also an inaccurate interpretation of Duhem's doctrine; Duhem's lack of direct support for the falsifiability thesis, a feature of Duhem's exposition acknowledged by Quinn, should warn us that the interpretation is dubious.

---

[1] Although non-separability (and non-fasifiability) might have been better designations. To simplify the discussion, one can regard the separability thesis as 'no theoretical hypothesis of physics by itself has observational consequences' and the falsifiability thesis as 'no theoretical hypothesis of physics can be conclusively falsified by observations alone'. However, the text shows that Duhem does not need so strong a thesis—'few or no theoretical hypotheses' will do in both cases. It might have been thought that the Duhem–Quine thesis is a conjunction of the separability and falsifiability theses; it should now be clear that the Duhem–Quine subthesis (i) in itself contains both the separability and falsifiability theses and that the Duhem–Quine subthesis (ii) is entirely independent.

There is another text that can help decide this question. Shortly before his death, Duhem composed a detailed overview of all his scholarly works in a document supporting his candidacy for membership to the Académie des Sciences. The 'Notice sur les titres et travaux scientifiques de Pierre Duhem' was a work devoted to describing Duhem's master scientific idea which dominated his thought throughout his life, the creation of a science that would unite the principles of mechanics and those of thermodynamics: energetics. The salient feature of energetics was that 'it did not imitate the numerous mechanical theories proposed at the time by the physicists [*i.e.* Maxwell's theories]; it did not substitute the hidden movements of hypothetical bodies for the observable properties that instruments measured' (Duhem [1917], p. 74). What energetics offered was a formal theory, with the character of a logical system, which instead of reducing physical qualities in the manner of mechanistic theories, limited itself to marking by means of a numerical scale the various intensities of such qualities. Duhem's critique of mechanistic theories (or the method of Cartesians and atomists) was that they are not autonomous:

> The physicist who wishes to follow them can no longer use the methods proper to physics exclusively [. . .]; he is carried into the domain of cosmology. He no longer has the right to shut his ears to what metaphysics wishes to teach him with respect to the real nature of matter. Hence physics becomes dependent on metaphysics and subject to the fluctuations of its doctrines. Thus the theories constructed according to the Cartesian or atomist method are condemned to infinite multiplicity as well as perpetual new beginnings; they appear far from being able to insure the general consent and continuous progress of science (ibid. [1917], p. 152; *cf.* also [1892] and [1954], pt. I, chap. i).

Duhem also contrasts the methodology of energetics with that of another great historical scientific methodology—the inductive method (or the Newtonian method)—which rejects any hypotheses about imperceptible bodies and hidden movements, and admits only general laws known by induction from observation. It is with respect to this contrast that the Duhemian thesis makes its first appearance. Duhem accuses the Newtonian method of being unmanageable ([1917], p. 152; *cf.* also [1894] and [1954], pt. II, chap. vi); he asserts that a science can only follow the Newtonian method when its means of knowing are those of common sense: 'Induction can no longer be practiced in this way when science no longer observes facts directly, but substitutes for them measurements given by instruments, of magnitudes that only a mathematical theory has defined' ([1917], p. 153). Duhem clearly indicates that some sciences can use the Newtonian method, derive their principles by induction from observation and disconfirm them as tests warrant. What Duhem thinks is wrong with inductivism is that it cannot be applied to all sciences.

According to Duhem, physics as a science has reached a stage in which the observational consequences of its theoretical hypotheses need to be interpreted by another chunk of theory; as a result, no theoretical hypothesis by itself has any observational consequences: 'a physics experiment is not simply the observation of a phenomenon. . . . It is the precise observation of a group of phenomena, accompanied by the *interpretation* of these phenomena; this interpretation substitutes for the concrete given, actually gathered by observation, some abstract and symbolic representations which correspond to the given by virtue of the physical theories admitted by the observer' ([1917], p. 153; *cf.* also [1954], p. 147). This is the separability thesis. Duhem's reason for it rests on his empirical thesis about a peculiar feature of experiments in physics. A physics experiment might require the measurement of the electrical resistance of a coil; another might require the measurement of the volume and temperature of a gas. These measurements cannot be observed directly; they require sophisticated instruments that rely upon other theories for their construction and operation, and for the interpretation of their results ([1894] and [1954], pt. II, chap. iv).

Having asserted the separability thesis, Duhem then states that there are consequences to be derived from this thesis, and proceeds to assert the falsifiability thesis: 'The physicist can never submit an isolated [theoretical] hypothesis to the controls of experiment, but only a whole set of hypotheses; when the experiment disagrees with his prediction, he learns that at least one of the hypotheses that constitutes the set is erroneous and must be modified' ([1917], p. 153 and [1954], p. 187). It is clear that Duhem is claiming that the falsifiability thesis is a result of the separability thesis, and that the separability thesis is an empirical thesis depending upon factors that do not govern all sciences. This reading is also supported by the fact that Duhem claims, in *La Théorie physique*, that he is imposing narrow limits on his inquiry, and is restricting his generalizations to physics ([1954], p. 3); moreover, when he first announces the separability thesis, he announces it as a principle about what physicists do, whose consequences will be developed in the remainder of the book ([1954], p. 147). One should note that this interpretation of Duhem's thesis, that it is an empirical thesis whose scope is restricted to physical science, a thesis aimed against inductivism (Newtonianism), does not fall prey to a paradox that lurks about in these matters. It would be paradoxical if Duhem were to hold that he has empirically refuted inductivism using a thesis that there can be no empirical refutations; Duhem can think that he has refuted inductivism empirically because he thinks that inductivism does not account for what physicists do (and perhaps it fails to account only for what physicists do).

3 SOME CONSEQUENCES OF FAILING TO UNDERSTAND DUHEM'S THESIS

Failure to understand that Duhem considers the falsifiability thesis as a consequence of the separability thesis can lead to confusions about what would count as a refutation of Duhem's position. For example, Popper argues against the Duhem thesis that (1) there is no reason why a counter-example to an axiomatised theoretical system may not be found to satisfy all of one's axioms except for one whose independence would thereby be established ([1963], pp. 238–9); (2) we can make use of our background knowledge in the search for a counter-example, and refute one or another theory rather than our background knowledge ([1963], pp. 239–40);[1] and (3) some scientific behaviour cannot be explained by one who adopts Duhem's view of scientific theorising ([1963], pp. 242–4; this is Koertge's main criticism—*cf.* [1978], p. 259).

It is clear that Popper's first argument fails against Duhem, since the Duhem thesis does not concern axiomatised theoretical systems, but whole theories, including the chunk of theory connecting the theoretical hypothesis to observation. It should be recalled that Duhem has formulated his thesis as a response to inductivism, and does not think it applicable to such methodologies as Cartesianism, in which the aim is to produce internally consistent axiomatised theoretical systems, which are externally consistent with a specific metaphysics. As long as Popper's argument attempts to establish the independence of an axiom, using a model to be constructed as a counter-example, the argument is not relevant to Duhem's point, given that Duhem's point is restricted to experimental physics. So Duhem may very well agree with Popper's statement, but disagree with it as a criticism.

If one continues to keep Duhem's objectives clear, he would see that Popper's second and third arguments must also fail. Again, Duhem announces his separability thesis against the view that inductivism is the only scientific methodology. He first establishes that it is false that all scientific statements confront observation directly, by arguing for the separability thesis that some scientific hypotheses—those of theoretical physics—are such that, by themselves, they cannot confront observation directly. Duhem does not claim that no statement of science *qua* science ever confronts observation directly. He merely claims that if the theoretical hypothesis requires other interpretive statements in order to confront experience (as hypotheses of physics apparently do), then observation is not decisive by itself (that is, the falsifiability thesis results). Duhem does not claim that scientists cannot decide such cases. He merely claims that such disconfirmations (or confirmations) are not conclusive. For some statements of science, background knowledge can be used to refute one or another theory; for the hypotheses of theoretical physics, if the background knowledge must be used in order to interpret the statement so that it may confront observation, then the background knowledge cannot escape the negative results of the observation. In that case, a scientist may still hold on to the background knowledge, and disconfirm the theory, but the disconfirmation would not be conclusive.

---

[1] In some respects, Clark Glymour's criticism of Duhem is just a reworking of this part of Popper's criticism; *cf.* Clark Glymour, [1975] or chap. 5 of [1980]. It should be noted that Quinn's objection (see above) is also an extension of this criticism—if we can verify a hypothesis, we can use it to refute some portion of theory.

Extensions of Popper's criticisms are numerous; it is often difficult to know against what the criticism is directed. For instance, Noretta Koertge begins her article, 'Towards a New Theory of Scientific Inquiry', in which the Duhemian problem is discussed and 'solved', with some remarks about common sense (as it is displayed in *Zen and the Art of Motorcycle Maintenance*, *i.e.*, in case of ignition failure, test the spark plugs before dismantling the carburettor—[1978], p. 253; *cf.* pp. 255–67 for the solution to the Duhemian problem). Koertge intends 'to begin to construct a theory of scientific inquiry that is at least as sophisticated and right-minded as that found in *Zen and the Art of Motorcycle Maintenance*' (*ibid.*, [1978]). But Koertge does not really discuss and solve the Duhemian thesis; at one level she accepts it, and at another she merely denies it. Koertge accepts that it is sometimes possible that a scientist confronted with a negative experimental result would reject the auxiliary assumptions of the test instead of the theory itself (that would be to accept Duhem's thesis). However, she thinks that there are decision-theoretical reasons for proceeding one way or another. Koertge acts as if one can always separate one's options neatly. But this is what Duhem denies. The foundation of the Duhem thesis is the impossibility to separate hypotheses for testing. Duhem does not deny that one can, at times, have good reasons for proceeding one way rather than another; he denies that these reasons can be conclusive. Duhem has even taken the trouble to reject the metaphors that would liken the physicist with the motorcycle repairman and his machine:

> Physics is not a machine that can be taken apart; one cannot try each piece in isolation and wait, in order to adjust it, until its solidity has been checked carefully. Physical science is a system that must be taken as a whole; it is an organism in which one part cannot be made to function without the more remote parts coming into play, some more than others, but all to some degree. If some discomfort, some illness is felt in the functioning, the physicist will have to guess, through the effect produced on the whole system, which organ needs to be remedied or modified, without being able to isolate this organ and examine it apart. The watchmaker to whom one gives a watch that does not function separates all the wheels and examines them one by one until he finds the one which is defective or broken. The doctor to whom a patient appears cannot dissect him in order to establish his diagnosis; he has to guess at the seat and cause of the ailment solely by inspecting disorders affecting the whole body. The physicist concerned with remedying a defective theory resembles the doctor and not the watchmaker ([1954], pp. 187–8).

4 FINAL COMMENTS ON THE DUHEM THESIS

It would seem that Duhem's thesis is so formulated that it escapes refutation by contemporary criticism (at least the criticism examined above). Duhem's thesis is sufficiently different from the Duhem–Quine thesis that it is not open to the charge that it is a *non-sequitur* to think that from separability and falsifiability we can conclude that scientists can hold any hypothesis come what may. Moreover, Duhem's primary thesis is separability, which he thinks is established empirically, and from which the falsifiability thesis derives. Duhem does not establish the separability thesis for all sciences, but merely for those sciences (theoretical physics, being the only example given) that require an interpretive chunk of theory in order for their theoretical hypotheses to confront observation. This engenders another distinction between Duhem's thesis and the Duhem–Quine thesis. Duhem thinks that theoretical physics has reached a stage such that its hypotheses do not have observational consequences; since this is an empirical thesis, Duhem would be willing to go back down from the universal thesis to a more modest thesis, if evidence can be presented against the universal thesis. The same cannot be said about the Duhem–Quine thesis. Given that the Duhem–Quine thesis concerns all statements and all evidence, it is not clear what can count as evidence against it such that its proponents would be willing to modify their views. In any case, Duhem's thesis is restricted both in the strength and in the scope of its conclusion. The separability thesis is directed solely against Newtonian inductivism; Duhem does not think that the argument can succeed against what he calls mechanistic theories (Cartesianism or atomism), although he thinks he has another argument against these theories—the great consensus in science, which allegedly could not be achieved if scientific theories were not independent from particular metaphysics (for which there is obviously no consensus). If one pays attention to these features of Duhem's argument one need not worry about the alleged contemporary refutations of Duhem's thesis. Of course, that is not to say that none of the attempted refutations have any merit—it is clear, for instance, that Grünbaum's argument succeeds against the broader interpretations of the Duhem thesis. Philosophers of science may legitimately wish to refute claims made for the Duhem thesis, but they should be careful to deny only that which is within the scope of their arguments.

Duhem is to be praised for setting forth an interesting thesis which attempts to navigate between the Scylla of 'reductionism' and the Charybdis of relativism.[1]

Virginia Polytechnic Institute and State University

---

[1] I have profited from comments and suggestions by members of Virginia Polytechnic Institute and State University's Department of Philosophy and Center for the Study of Science in Society.

REFERENCES

BLOOR, D. [1976]: *Knowledge and Social Imagery*. Routledge and Kegan Paul.

BLOOR, D. [1981]: 'The Strengths of the Strong Programme', *Philosophy and Social Criticism*, **11**, pp. 199–213.

DUHEM, P. [1892]: 'Quelques réflexions au sujet des théories physiques', *Revue des Questions scientifiques*, 2nd series, **1**, pp. 139–77.

DUHEM, P. [1893a]: 'Physique et Métaphysique', *Revue des Questions scientifiques*, 2nd series, **2**, pp. 55–83.

DUHEM, P. [1893b]: 'L'école anglaise et les théories physiques', *Revue des Questions scientifiques*, 2nd series, **2**, pp. 345–78.

DUHEM, P. [1894]: 'Quelques réflexions au sujet de la physique expérimentale', *Revue des Questions scientifiques*, 2nd series, **3**, pp. 179–229.

DUHEM, P. [1904–5]: 'La théorie physique, son objet et sa structure', *Revue de Philosophie*, 4th year, vol. 1, pp. 387–402, 542–56, 643–71; vol. 2, pp. 121–60, 241–63, 353–69, 535–62, 712–37, and 5th year, vol. 1, pp. 25–43, 267–92, 377–99, 519–59, 619–41.

DUHEM, P. [1905–6]: 'Physique de croyant', *Annales de Philosophie chrétienne*, 77th year vol. 1, pp. 44–67 and 133–59. Included as an appendix in Duhem [1914] and translated in Duhem [1954].

DUHEM, P. [1906]: *La Théorie physique, son objet et sa structure*. Chevalier et Rivière.

DUHEM, P. [1908]: 'La valeur de la théorie physique, à propos d'un livre récent', *Revue Générale des sciences pures et appliquées*, **19**, pp. 7–19. Included as an appendix in Duhem [1914] and translated in Duhem [1954].

DUHEM, P. [1912]: 'La nature du raisonnement mathématique', *Revue de Philosophie*, 12th year, pp. 531–43.

DUHEM, P. [1913]: 'Examen logique de la théorie physique', *Revue scientifique*, 51st year, pp. 737–40.

DUHEM, P. [1914]: *La Théorie physique, son object et sa structure*. 2nd enlarged ed. Rivière. Reissued, Vrin, 1981.

DUHEM, P. [1915]: *La Science allemande*. Hermann.

DUHEM, P. [1917]: 'Notice sur les titres et travaux scientifiques de Pierre Duhem', *Mémoires de la Société des Sciences physiques et naturelles de Bordeaux*, pp. 71–169.

DUHEM, P. [1954]: *The Aim and Structure of Physical Theory*. Translated P. Wiener. Princeton University Press, Paperback edition, Atheneum Press, 1962. All page references to Duhem [1954] will be to the 1962 paperback.

FRANK, P. [1941]: *Modern Philosophy and Its Science*. Harvard University Press.

GLYMOUR, C. [1975]: 'Relevant Evidence', *The Journal of Philosophy*, **72**, pp. 403–26.

GLYMOUR, C. [1980]: *Theory and Evidence*. Princeton University Press.

GRÜNBAUM, A. [1960]: 'The Duhemian Argument', *Philosophy of Science*, **27**, pp. 75–87.

HESSE, M. [1980]: *Revolutions and Reconstructions in the Philosophy of Science*. Indiana University Press.

KNORR-CETINA, K. D. and MULKAY, M. [1983]: 'Introduction: Emerging Principles in Social Studies of Science', in K. D. Knorr-Cetina and M. Mulkay (*eds.*): *Science Observed: Perspectives on the Social Study of Science*. Sage Publications.

KOERTGE, N. [1978]: 'Towards a New Theory of Scientific Inquiry', in G. Radnitsky and G. Andersson (*eds.*): *Progress and Rationality in Science*. Reidel.

LAKATOS, I. [1970]: 'Falsification and the Methodology of Scientific Research Programmes', in I. Lakatos and A. Musgrave (*eds.*): *Criticism and the Growth of Knowledge*. Cambridge University Press.

LAUDAN, L. [1965]: 'Grünbaum on the Duhemian Problem', *Philosophy of Science*, **32**, pp. 295–9.

LAUDAN, L. [1982]: 'Overestimating Underdetermination', Philosophy Colloquium, Virginia Polytechnic Institute and State University.

LOWINGER, A. [1941]: *The Methodology of Pierre Duhem*. Columbia University Press.

MARTIN, R. N. D. [1981]: *The Philosophy of Physics According to Pierre Duhem*. Unpublished Dissertation. London School of Economics.

MARTIN, R. N. D. [1982]: 'Darwin and Duhem', *History of Science*, **20**, pt. 1, n. 47, pp. 64–74.

PICKERING, A. [1980]: 'The Role of Interests in High-Energy Physics', in K. D. Knorr, R. Krohn, and R. Whitley (*eds.*): *The Social Process of Scientific Investigation*. Reidel.

POPPER, K. [1963]: *Conjectures and Refutations*. Routledge and Kegan Paul.
QUINE, W. V. O. [1953]: 'Two Dogmas of Empiricism', in *From a Logical Point of View*. Harvard University Press.
QUINE, W. V. O. [1969]: 'Epistemology Naturalized', in *Ontological Relativity and Other Essays*. Columbia University Press.
QUINE, W. V. O. [1976]: 'A Comment on Grünbaum's Claim', in S. Harding (*ed.*): *Can Theories be Refuted?* Reidel.
QUINN, P. [1974]: 'What Duhem Really Meant', in R. S. Cohen and M. W. Wartofsky (*eds.*): *Methodological and Historical Essays in the Natural and Social Sciences*. Reidel.
QUINN, P. [1978]: 'Rejoinder to Tuana', *Philosophy of Science*, **45**, pp. 463–5.
TUANA, N. [1978]: 'Quinn on Duhem: An Emendation', *Philosophy of Science*, **45**, pp. 456–62.

Brit. J. Phil. Sci. 35 (1984), 327–343 *Printed in Great Britain*

The Passivity Assumption of the Sensation–Perception Distinction*

by AARON BEN-ZEEV

The sensation–perception distinction did not appear before the seventeenth century, but since then various formulations of it have gained wide acceptance. This is not an historical accident and the article suggests an explanation for its appearance. Section 1 describes a basic assumption underlying the sensation–perception distinction, to wit, the postulation of a pure sensory stage—*viz*. sensation—devoid of active influence of the agent's cognitive, emotional, and evaluative frameworks. These frameworks are passive in that stage. I call this postulation the passivity assumption. Section 2 suggests three major reasons for the emergence of this assumption in the seventeenth century: the mental–physical gap, the causal theory of perception, and epistemological considerations regarding the status of the sensory given. In the last section a critical discussion is presented. The passivity assumption is found to have serious empirical and theoretical flaws.

1 *Description of the Passivity Assumption*
2 *The Reasons for the Emergence of the Passivity Assumption*
3 *Criticism of the Passivity Assumption*

1 DESCRIPTION OF THE PASSIVITY ASSUMPTION

In its most prevalent formulation, the sensation–perception distinction is the distinction between two mental stages: one devoid of active mental contributions—sensation—and the other, the result of the former stage and active mental contributions—perception. We may therefore call the approach that assumes this distinction The Dualist Approach to Perception. A basic feature of the above distinction is the passivity assumption, *i.e.*, the postulation of a sensory stage in which the mind is passive. This is true with regard to cognitive, emotional and evaluative contributions of the agent. Sensation is purified from such contributions. The emphasis is on cognitive purity. Our discussion of this approach will concentrate on the cognitive aspect; the discussion can, however, also be applied to the assumed emotional and evaluative purity. (For a more detailed description of the dualistic approach see Ben-Zeev and Strauss [1984]).

In the dualistic approach, sensation is the passive reception of sensory

Received 8 August 1983

---

* I am grateful to Stephen Toulmin, Alan Gewirth, Robert Richards, Michael Strauss, and William Wimsatt for helpful discussions. While I began the work on this article I enjoyed a grant from the Rothschild Foundation.

input, whereas perception includes active cognitive elaborations of that input. While sensations are the immediate correlate (or the inevitable result) of physical states, perceptions are derived from sensations and supplementary cognitive processes. The most common (though not indisputable) examples of sensations are secondary qualities, *e.g.*, qualities such as colour and sound, and feelings such as pain and thirst. These are regarded as merely starting-point and not a final product of cognitive processes. Examples of perception include primary qualities such as distance and movement and meaningful objects (chairs and trees) and events (*e.g.*, that the stone breaks the window).

We may describe the sensation–perception distinction in light of the following five traditional factors assumed to be involved in the causal process of perception: (*a*) the stimulus' source (what is usually called 'the distal stimulus'), *e.g.*, the table; (*b*) the stimulus (or in its traditional term 'the proximal stimulus'), *e.g.*, the light rays coming from the table; (*c*) physiological processes; (*d*) sensation; and (*e*) perception. The dualistic approach may then be formulated as follows:

1. The relationship between the stimulus' source and the stimulus is equivocal; to wit, the latter does not unequivocally specify the former.
2. Sensation does not display any degree of independence from the stimulus or its inevitable physiological effect; therefore, sensation does not involve active mental contributions.
3. Hence, sensation cannot reflect the stimulus' source.
4. Perception is not in correspondence with the stimulus, but it is supposed to reflect the stimulus' source.
5. Hence, perception must involve cognitive activities which overcome the equivocation of the stimulus.

For our present purposes the second premise is the most important; it expresses the passivity assumption. Sensation does not display any degree of independence from the stimulus or its inevitable physiological effect since sensation itself is the inevitable effect of those factors. Therefore, we do not need to assume the participation of the cognitive (as well as the emotional and evaluative) system in the creation of that stage. In order to sustain that position, one has to show that the typical features of cognition are indeed absent from that stage. A central feature of cognition is the establishment of certain relations and order (where order is an organized complex of relations). Indeed, the establishment (and even, to a certain extent, the reception) of relations is denied from the stage of sensation. This is true regarding all kinds of cognitive factors: memory (which draws relations to past experience), imagination (relations to something which is not present), judgement (relations to one's knowledge), expectation (relations to assumed future events), *etc.*

In the light of the passivity assumption we can understand Russell's contention that 'the essence of sensation . . . is its independence of past experience'; and in order 'to arrive at what really is sensation . . . we have to pare away all that is due to habit or expectation or interpretation' (Russell [1921], pp. 144, 140). Sensation is then what remains constant through all the changes that the cognitive agent has undergone due to past experience or (cognitive) maturation; it contains no relation to his past experience. A sense-datum is 'the core, in a perception, which is solely due to the stimulus and the sense-organ, not to past experience' (Russell [1927], p. 204). Hence, sensations 'have not the characteristics that we associate with the word "cognition". . . . Thus while the sensory core is not cognitive, its associative accompaniments, being beliefs, must be classed as cognitions' (Russell [1940], pp. 119–21). Therefore, Russell asserts:

> But although sensation is a *source* of knowledge, it is not itself, in any usual sense, knowledge. . . . 'Perception', as the word is used by most psychologists, is of the nature of knowledge, but it is so because of the adjuncts which are added to pure sensation by experience, or possibly, by congenital dispositions ([1948], p. 422).

The above contentions can be found throughout the philosophical and psychological literature since the seventeenth century. Here are a few examples. The passive stage of sensation that I mentioned is what Descartes terms 'the second grade of sensation'. This grade 'is due to the fact that the mind is so intimately conjoined with the brain as to be affected by motions arising in it. Nothing more than this should be assigned to sense, if we wish to distinguish it accurately from the intellect' ([1641], p. 252). In Locke's theory, the passive stage is the stage of sensing simple ideas. Regarding this stage 'the understanding is merely passive; and whether or not it will have these beginnings, and as it were materials of knowledge, is not in its own power' ([1690], II, 1, 25). In accordance with the passivity assumption, Berkeley argues that 'To perceive is one thing; to judge is another. So likewise, to be suggested is one thing, and to be inferred another. Things are suggested and perceived by sense. We make judgements and inferences by the understanding' ([1733], sec. 42). The article on the senses in Diderot's *Encyclopedia* also states the passivity assumption: 'The senses always communicate faithfully the impressions they receive. This is obvious because the senses are natural faculties dependent on the impressions they receive from objects' ([1965], p. 329). Likewise, Halleck argues that

> perception is that power which interprets the raw materials given by sensations. Perception, always aided by the other faculties, gives us our first exact knowledge. . . . To our mental ear sensations are all the while speaking, but their language is unintelligible until perception interprets it ([1895], pp. 66–7).

And in a later psychology textbook we find the same view: 'Sensations are "pure" experiences in that they do not have organization, meaning, or association. Perception is conscious experience that has some degree of organization and meaning' (Shaffer *et al.* [1940], p. 202). In accordance with the passivity assumption which states that the intellect or factors of past experience are not involved in sensation, Helmholtz argues that 'no indubitable present sensation can be abolished and overcome by an act of the intellect' or by 'factors of experience' ([1867], pp. 12–13). C. I. Lewis provides another formulation of the passivity assumption: 'There are, in our cognitive experience, two elements: the immediate data, such as those of sense, which are presented or given to the mind, and a form, construction, or interpretation, which represents the activity of thought' ([1929], p. 38). And more recently, Dretske claims that

> if one systematically strips away from a given perceptual act all the accretions due to past experience, all the collateral information, anticipations, interpretive and inferential elements, all the habitual or conditional associations, then one will be left with a "pure sensory core"—the *given of sense experience* ([1969], p. 75).

It should be noted that the passivity assumption refers to the mental (mainly cognitive) factor of the agent and not necessarily to other factors. Thus the physiological system may be active in the sense that it is the origin of its own motion or the motion of other objects. Descartes, for instance, asserts that:

> all our external senses, in so far as they are part of the body, and despite the fact that we direct them towards objects, so manifesting activity, *viz.* a movement in space, nevertheless properly speaking perceive in virtue of passivity alone, just in the way that wax receives an impression from a seal ([1628], p. 36).

Similarly, in Berkeley's view, minds are active in the metaphysical realm—namely, they are the origin of the existence of everything—whereas in the psychological realm there is a sensory stage (*e.g.*, sensing the 'immediate objects of vision') in which the cognitive, emotional and evaluative factors of the mind are passive.

Another clarification regarding the passivity assumption concerns the notion of active and passive mental contributions. I have claimed that sensation is devoid of active mental contributions, but not of passive ones. Thus, for example, sensation is usually assumed to consist of secondary qualities such as colour and pain which are regarded as mental *i.e.*, containing some (passive) mental contributions. In that view secondary qualities are mental qualities that do not exist in the 'real' physical world. Their creation involves mental contributions since they are products of an interaction of which one factor is mental. However, those are passive contributions since the mind cannot change these qualities at will; they are inevitable mental expressions of previous physical states. Mental 'stuff' is one factor which is responsible for the creation of sensation, but the individuation of that stuff (*i.e.*, the different personalities and cognitive structures of the agent) does not influence the sensory content. Therefore, the typical mental faculties—namely the cognitive, emotional and evaluative faculties—are passive in that stage. Those faculties are active only in the creation of the later perceptual stage.

2 THE REASONS FOR THE EMERGENCE OF THE PASSIVITY ASSUMPTION

A detailed explanation of the emergence in the seventeenth century of the sensation–perception distinction and its passivity assumption requires a description of the different conceptual frameworks of ancient and modern science and philosophy. I do not intend to assume this task here. I will merely indicate three closely connected reasons for this emergence: the mental-physical gap, the causal theory of perception, and epistemological considerations regarding the status of the sensory given.

The mental-physical gap, which has been a central issue in philosophy since the seventeenth century, has proved to be a problematic feature in every theory of perception formulated since. This is due to the fact that sensation is regarded as a mental event which is the immediate correlate (or effect) of physical events. The mental-physical gap, however, is not as large in ancient theories. Thus Aristotle assumes a kind of continuity, and not a gap, between psychical and physical processes. We can find, according to Aristotle, psychical processes not only in human beings but also in animal and plants: 'the principle found in plants is also a kind of soul' (*De Anima* 411 b28). In this way the psychical–physical gap is smaller, as the gap between plants and inorganic bodies is much narrower than the modern gap between conscious, mental deliberation and inorganic bodies. The main distinction that Aristotle makes is not between mind and body, but between living and non-living; the gap in the latter distinction is smaller than the former. Also in Plato's theory, which assumes different kinds of existence for different entities, the mental-physical distinction does not play a crucial role. Plato regards as mere appearances both sensory qualities ('that which appears') and physical objects to which they refer ('that which is always in a process of becoming and perishing'). By the same token, what really exist are 'the very things themselves . . . ever remaining the same and unchanged', or the 'self-existent ideas unperceived by sense' (*Republic* 478–80; *Timaeus* 27–8, 51–2). The ancient view avoids a clear-cut dichotomy between the human mind and the 'rest of the world'. Therefore, Toulmin claims that while the Greek scientists did not separate the known object from the knowing subject, the seventeenth-century scientists did:

> in the Ancient's view, human beings formed part of a single comprehensive *cosmos* along with physical and biological Nature. . . . The Ancients found nothing particularly surprising about the intelligibility of Nature: by their nature, the human reason and the universal *logos* has a close affinity or analogy. . . . [On the other hand, the] self-styled "new philosophers" of the seventeenth century . . . saw human mentality and physical nature as belonging to distinct realms. . . . The rational operations of the human mind, seemingly, had nothing in common with the causal workings of the physical world ([1981]; see also Smith [1974]).

In the perceptual realm, the mind-body dichotomy of the seventeenth century creates a qualitative gap between the 'metaphysical matter' of sensation (and perception) and that of the external, 'real' world; the former being mental while the latter is physical. The mind in this view cannot perceive objects which are not mental. Therefore, there is no direct perception of external objects: 'the human mind does not perceive any external body as actually existing, except through the ideas of the modifications of its own body' (Spinoza [1677], II, 26; see also Berkeley [1713], III, pp. 188, 207). The perception of external objects is indirect here: it is achieved through the reception of ideas that represents the objects. (T. Reid presents a notable exception to this approach.)

The qualitative gap between the perceptual and physical stuff exists neither in the common sense view nor in the ancient one. In the common sense view there is an identity between perceptual objects and physical things. One perceives real trees and chairs and not merely mental representations of them. In Greek philosophy the identity is reduced to a certain part of the perceptual content, *viz.* the form. Aristotle, for example, believed that the same perceptual form existed in physical objects as in the perceiving mind; only the matter by which they were individuated differed. Similarly, while colour is conceived by most modern perceptual theories as merely a mental subjective sign which does not exist in 'real' physical objects, in almost all (if not all) ancient theories of vision, colour is a real property of bodies (see Lindberg [1976], pp. 1–17). Cassirer characterises the essential functions of the Greek manner of thinking as 'merely those of comparing and differentiating a sensuously given manifold'. The generic concept which is basic to this way of thinking 'does not appear as something foreign to sensuous reality, but forms a *part* of this reality; it is a selection from what is immediately contained in it' ([1910], p. 5.[1]

The passivity assumption was, in a sense, one way of bridging the mind–body gap in the perceptual realm. Thus, while sensations are determined by physical events, they already exhibit some primitive mental attributes. These attributes are mental but they do not require the typical active and complex operations of the mind. Hence, they can be regarded as mediators between the physical and mental realms. It is natural to assume that the mental attributes which are absent in that intermediate stage are those that seem to consist of higher mental activities, *i.e.*, cognitive activities which characterise conscious rational thinking. If these complicated mental activities were found in sensation, sensations had to be typical mental entities and not an intermediate stage between the mental and physical. Indeed, the second grade of sensation in Descartes' view (what we simply call 'sensation') is an intermediate stage between the physiological grade (the first grade) and the typical mental one (the third grade). The role of sensation as an intermediate stage, as a 'boundary between the mental and the physical', is evident in Russell's position: 'Sensations are what is common to the mental and physical worlds; they may be defined as *the intersection of mind and matter*.' And 'those that have physical causes and mental effects we should define as "sensations"' ([1921], pp. 144, 138).

---

[1] Cassirer points out, as the basic reason for the differences in Greek and modern periods, the transition from a generic concept (*Substanzbegriff*) to the concept of mathematical function (*Funktionsbegriff*) as a basic conceptual framework. That is, the transition, in the basic scientific explanatory unit, from genus and species to mathematical relations which exist in mathematical functions. In his explanation of the above differences, Toulmin [in press] emphasises the different concepts of matter in the two periods; in the seventeenth century, matter was usually conceived of as inert and passive while in Greek times, matter was endowed with active elements. The different emphases reflect the fact that natural philosophy of the seventeenth century was primarily a mathematical and mechanistic science. Toulmin suggests that in our description of that view we should stress the mechanistic side, since the mental–physical gap is absent in the few modern theories which do not adopt the passive, mechanistic depiction of matter. Thus, indeed, the gap does not exist in Priestley's view which rejects the assumption that matter is 'impenetrable, inert substance'. He claims that 'matter, destitute of what has hitherto been called solidity, being no more incompatible with sensation and thought than that substance which without knowing anything farther about it, we have been used to call immaterial' (Priestley [1777], p. 111).

A second basic reason for postulating that the stage of pure sensation, which is devoid of cognitive contributions, is connected to the causal theory of perception. This theory assumes a unidirectional causal chain beginning with the objective physical world and ending with subjective percepts: 'Percepts are in our heads, for they come at the end of a causal chain of physical events leading, spatially, from the object to the brain of the percipient' (Russell, [1927], p. 320). In the visual domain this chain is roughly as follows: light reflected from an object enters the eye; a retinal image is created in the eye; the image is transferred through the optic nerve to the brain; brain activity takes place; and sensation of patches of colour occurs (see, *e.g.*, Descartes [1628], pp. 36–40). Hence, sensations 'are subjective effects of objective processes'; and, 'the *rise of sensations* . . . is regularly dependent on certain physical processes that have their origin partly in the external world surrounding us, partly in certain bodily organs' (Wundt [1907], pp. 5, 41).

There is a dispute as to whether the mental sensation is the last member of the causal sequence or merely the immediate correlate (and not the direct effect) or the last physical event in the sequence. This dispute stems from the difficulty in explaining the derivation of mental effects from physical causes. Thus, though Wundt conceives of sensation as the direct effect of physical and physiological stimuli, he admits that 'it is impossible to explain the character of sensation from the character of physical and physiological stimuli. Stimuli and sensations can not be compared with one another at all' ([1907], p. 49).

An underlying principle in causal theories is that the effect cannot be greater than its cause, *i.e.*, there are no properties of the effect which were not, somehow, in the cause (otherwise something would be produced out of nothing). This principle—as well as other considerations—raises the problem of how mental properties can be the effects of physical causes. No matter what the solution is, the problem is less severe when the effect manifests at a very primitive mental stage, *viz.* a sensory core which is devoid of complex mental contributions (such as the cognitive ones). The value of the passivity assumption in dealing with this problem is obvious. The assumption is also useful in the version which conceives of a sensation as the immediate correlation (and not the direct effect) of physiological excitations in the brain. Properties of external objects may somehow be represented in the brain, but, strictly speaking, they are not in the brain: the brain is thought to contain properties of the (proximal) stimulus, but not of the stimulus' source. Therefore, a distinction is needed between the mental stage which merely reflects the (proximal) stimulus' properties, *viz.* the impoverished stage of sensation, and the mental stage which reflects the properties of the stimulus' source, *viz.* the rich stage of perception. The causal theory of perception was another reason for the emergence of the sensation–perception distinction and its passivity assumption.

As a result of the intermediary function of sensation in the mind-body gap and the causal theory of perception, another perceptual gap was created: a quantitative gap between properties represented in sensations (or those that exist in the stimulus) and properties of external objects (the stimulus' source) (see also Shaw *et al.* [1981]). Since sensations do not exhibit many attributes they cannot represent the richness and complexity of the objects' properties. At most, sensations correspond with the stimuli (or merely—as Spinoza argues—with one's own bodily modifications), but never with external objects.

In light of the above qualitative and quantitative gaps, sensations were conceived as mental signs of physical objects; and a sign 'need not be similar in any way to that of which it is a sign' (Helmholtz [1878], p. 37). Indeed, since the seventeenth century science has conceived its subject matter as existing beyond the sensory qualities. These subjective qualities are merely signs of the objective, physical world: 'the point of view of natural science may . . . be possible only after abstracting from the subjective factor present in all actual experience' (Wundt [1907], p. 3). Planck states the same claim: 'The sense perceptions have been definitely eliminated from the physical acoustics, optics and heat. The physical definitions of sound, colour, and temperature are today in no way associated with the immediate perception of the respective senses' ([1925], p. 5).

I should note that the above quantitative gap was not common in ancient thinking. Thus, the discovery that three-dimensional attributes of objects are not represented in the sense organ in a direct or simple manner was not a trivial one:

> In several types of ancient visual theory, no difference was postulated between the spatial properties of objects and those of the optical stimulus, because it was believed that the stimulus to vision shares the attributes of the seen objects (Hatfield and Epstein [1979], p. 365).

Hence, it is no wonder that

> one question that has persistently plagued modern philosophy was never explicitly discussed by Aristotle, namely how and in what manner we perceive or come to be

aware of the concrete, physical objects to which sense qualities belong. Aristotle, it seems, never considered this an important philosophical question; he nowhere discusses it directly. No Greek, for that matter, ever addressed himself to this question (Block [1960], p. 93).

This question can arise only after the postulation of qualitative and quantitative gaps between appearances and physical objects. I do not wish to claim that in Greek philosophy there is no distinction between what appears to us and other entities. The pre-Socratic philosophers had already tried to point out the substance behind all appearances. However, these appearances were not regarded as *mental signs*; hence, the above gaps were absent. Concerning the mental–physical gap, I have shown that in ancient theories the borderline between what appears to us and the objective world is not that between mental and physical, rather that it crosses the mental–physical distinction. Concerning the second gap, I have claimed that, contrary to modern sensory signs, sensory appearances in ancient theories share many common properties with objects in the physical world.

The qualitative and quantitative gaps between pure sensations and the objects' properties create severe epistemological difficulties. These difficulties can be illustrated in Ryle's depiction of the modern approach to sensation:

> There is immured in a windowless cell a prisoner, who has lived there in solitary confinement since birth. All that comes to him from the outside world is flickers of light thrown upon his cell-walls and tappings heard through the stones; yet from these observed flashes and tappings he becomes, or seems to become, apprised of unobserved football-matches, flower-gardens and eclipses of the sun (Ryle [1949], p. 223).

Indeed, Locke asserts that the senses are

> the windows by which light is let into this *dark room, viz.* the understanding. For, me thinks, the understanding is not much unlike a closet wholly shut from light with only some little openings left, to let in the external visible resemblances, or ideas of things without ([1690], II, 11, 17).

(These comparisons are somewhat similar to Plato's cave image. However, while they refer to the gap between the sensory content and the real physical world, Plato's image refers to the gap between human understanding and the realm of Ideas.) Such difficulties prevent one from being satisfied with Aristotle's contention that in normal conditions the organism must have reliable perceptual knowledge of its environment (see Ben-Zeev [1984b]). Every modern theory of perception begins by questioning whether such knowledge is possible at all, and by seeking the foundations of perceptual knowledge. Because of the current enormous gaps between the starting point of cognitive activities and the (objective) world which the final cognitive products should depict, the starting point has created new epistemological dilemmas. If there is any hope for the mind to link to the objective world, it is in that pure starting-point of sensation. Hence, the notion of sensory given has become central.

Taking into consideration the aforementioned gaps, the necessity of sensory given is obvious: 'If there be no datum given to the mind, then knowledge must be contentless and arbitrary; there would be nothing which it must be true to' (Lewis [1929], pp. 38–9). In order to fulfil its epistemological task as the basis for cognitive elaborations, the given should have what seem to be two contradictory features: (a) it should be devoid of cognitive contributions; and (b) it should have cognitive content. The purpose of the first feature is to prevent a circular epistemological situation in which cognitive elaborations are confirmed by something which already involves cognitive contributions. The necessity of the second feature stems from the fact that contentless (meaningless) material cannot determine a correct cognitive elaboration since it has no relation to any of the possible cognitive elaborations. These two features appear contradictory as it is hard to conceive of any cognitive content which is devoid of cognitive contributions.

It is easy to see the role of the passivity assumption in connection with the above epistemological considerations. This assumption fulfils the first required feature of the given, and in a way it suggests some sort of solution for the second feature too. Sensations, we may recall, fully correspond with the (proximal) stimuli. Therefore, (a) sensations cannot be false (since physical events cannot be false), and (b) they derive their content from objective, and not subjective events. Consequently, sensations are governed by objective regularity which can be revealed by objective (physical and physiological) research. The absence of cognitive contributions in sensations, and hence their non-arbitrary way of creation, is the epistemological guarantee (of the dualistic approach) for revealing the objective content behind them. Helmholtz clearly expresses this epistemological consideration:

> Thus, even if in their qualities our sensations are only signs whose specific nature depends completely upon our makeup or organization, they are not to be discarded as empty appearances. They are still signs of something—something existing or something taking place—and given them we can determine the laws of these events. And that is something of the greatest importance ([1878], pp. 372–3).

In order to prevent any distortion of the sensory content, a passive cognitive agent, and often even a passive physiological conduit, was assumed. The sensory organs or the nerves are merely, as Locke asserts, 'the conduits to convey them [the ideas of one sense] from without to their audience in the brain—the mind's presence-room' ([1690], II, 3, 1). This conduit is passive (indifferent) in the sense that its properties do not influence the properties of the material transferred in it. Indeed, the fact that sensations are considered to be in full correspondence with the stimuli indicates that the physiological process which is between them, should be regarded merely as a passive transmission.

Köhler points out that such a view—which he terms 'the machine theory'—assumes that 'dynamic factors are completely prevented from influencing the distribution of the neural flow'; hence, 'the nervous system is devoid of any characteristic processes of its own'. A basic assumption behind this view is what Köhler calls 'the constancy hypothesis', *i.e.*, 'the point-to-point determination of local sensory facts by local stimuli ([1929], pp. 66–70). That is, the local stimuli alone determines, without the participation of the cognitive agent, the sensory qualities; hence, these qualities are 'pure'. Toulmin has a similar characterization for the above neurological model which he claims has been widely accepted since the seventeenth century; 'a sensory nerve was a material fiber through which some intermediate causal agency was transmitted linearly to the depths of the brain, where all mental processes of "perception" were actually initiated' ([1967], p. 825; Toulmin points out that the odd-men-out in that widely accepted model were Leibniz and Harvey both of whom retained a strong affinity with Aristotle). A similar model can be found—as Melzack testifies—in pain theories: 'specificity theory proposes that a specific pain system carries messages from pain receptors in the skin to a pain centre in the brain'; the model of such a theory is 'that of a fixed, direct-line communication system from the skin to the brain' ([1973], pp. 126, 134).

The above references clearly indicate an attempt to retain an objective origin to sensation. If indeed the processes preceding sensation do not distort its correspondence with the 'objective' stimulus, it may be regarded as the objective starting point of the whole edifice of knowledge. In light of the veil of ideas which mediates, in this view, the agent's knowledge of its environment, the existence of such an objective starting-point is of utmost importance. As mentioned before, the ancient theories of perception assumed that some actual property of the external object (*e.g.*, its form) is literally transferred into the perceiver. Since that form is the essence of the object, there is no problem of a veil covering the external objects. However, once this direct realism was abandoned, and the perceiver was conceived to be directly aware of mere mental entities, it was crucial to assume that, from an epistemological viewpoint, some aspects of these entities were not distorted.

I wish to note that in ancient theories the perceptual process had some passive aspects as well. Thus, the forms of the objects were not changed through that process. However, in ancient theories the perceptual passivity was not a cognitive one, and consequently it was not an epistemological bridge. As indicated above, only in modern, but not in ancient philosophy, there is an epistemological gap between the sensory given and external objects.

To summarize, I have delineated three closely connected reasons for the emergence, in the seventeenth century, of the passivity assumption which postulates a sensory stage devoid of any cognitive (as well as emotional and evaluative) contributions. The first reason is connected to the enormous gap—which emerged at that time—between the mental and physical realms. The sensory stage is an intermediate stage in the gap. Hence, it does not contain any active mental contributions such as the cognitive ones. The second reason is linked to the causal theory of perception. In the causal chain, which begins in the physical world and ends in mental sensations, the first sensory stage has a physical cause (or is the immediate correlate of such a cause). Hence, that stage cannot be the result of mental, cognitive activities which are found only in later stages of the causal chain. The third reason for the emergence of the passivity assumption stems from epistemological considerations connected with the modern gaps between sensations and the objects' properties. If the initial sensory stage—which is only a sign and not a picture of the external objects—already contains cognitive contributions which may be mistaken, then knowledge would be (as Lewis argues) 'contentless and arbitrary; there would be nothing which it must be true to'. In a sense, all these closely connected reasons stem from the contention that (as Russell puts it) 'sensations are what is common to mental and physical worlds; they may be defined as the intersection of mind and matter'. Sensations are considered primitive mental entities with physical causes. Their physical origin prevents them from being caused by complex mental activities (such as cognitive ones), and hence sensations cannot be false.

3 CRITICISM OF THE PASSIVITY ASSUMPTION

After presenting the passivity assumption of the sensation–perception distinction and the reasons for its emergence, I wish now to offer a critical evaluation of the view. There are two parts in this evaluation: (A) An examination of empirical evidence for the existence of the pure sensory stage (sensation) postulated by the passivity assumption; (B) A discussion regarding the theoretical merits of the assumption.

A central contention of the passivity assumption is that there is a sensory stage (sensation) which is devoid of cognitive, emotional and evaluative contributions and that this stage is qualitatively different from the subsequent stage of perception. I believe there is none whatsoever. Consequently the advocates of this view are neither certain nor in agreement regarding the empirical identification of that stage.

The initial and most common identification of pure sensation is with secondary qualities, *i.e.*, qualities such as colour, sound and smell, and feelings such as pain and pleasure. There is no empirical evidence for the assumption that secondary qualities are devoid of cognitive, emotional and evaluative contributions, and hence depend only on the stimulus and not on mental variables. Contrary to that assumption, empirical research indicates that there is no direct correspondence between perceiving a secondary quality and the stimulus' variables. Thus the constancy phenomenon also exists in colour perception. In many cases we perceive an object as having the same colour even if the illumination of that object (and hence the local proximal stimulus, *viz.* the light reflected from the object to the eyes) has changed. The same applies to the feeling of pain. Empirical evidence clearly suggests that 'the amount and quality of perceived pain are determined by many psychological variables in addition to the sensory input' (Melzack [1973], p. 136). Pain is determined by many personal, cognitive and cultural factors besides bodily image:

> The psychological evidence strongly supports the view of pain as a perceptual experience whose quality and intensity are influenced by the unique past history of the individual, by the meaning he gives to the pain-producing situation and by his "stage of mind" at the moment . . . In this way pain becomes a function of the whole individual, including his present thoughts and fears as well as his hopes for the future (Melzack [1973], p. 48).

Similar empirical conclusions have been reached in regard to all other perceptual qualities. And in general,

> The basic structuralist attempt to observe the sensations that presumably compose our perceptions of objects and events has failed for every physical attribute of objects that has been studied! We can list these failures under three headings: the *perceptual constancies*, the *illusions*, and the facts of *organization* (Hochberg [1978], p. 71).

The three phenomena mentioned as the major reasons for the failure to prove the existence of a stage of pure sensation clearly indicate the necessity of referring to contributions of the cognitive (as well as emotional and evaluative) system to explain perceptual (or sensory) experience.

Another widespread empirical focus of the stage of pure sensation has been that of infant perception. It has been claimed that the absence of past experience from infant perception makes it a good purveyor of pure sensation. This contention, which conceives infant perception to be an unorganised, meaningless flux ('one great blooming, buzzing confusion' in James' terms), has long prevailed since there were no satisfactory experimental techniques to test its validity. Current psychological research in neonate perception points out that the above contention is incorrect; the perceptual world of the neonate is complex and contains meaningful features typical of perception. Infants less than one month old demonstrate discrimination of visual patterns, colours, properties of sound, and tastes. Also, perceptual constancies and recognition of the mother occur at an early age. Accordingly, the abandonment of the assumption concerning the purity of infant perception is inevitable: 'the demonstration of fairly good visual acuity in the neonate and a wealth of other research on pattern perception, space perception and hearing has markedly changed our view of the young child' (Walk [1978]; see also Bower [1982]).

The rejection of the 'purity' of infant perception is further confirmed by empirical evidence concerning animal neonates. Early formulations of the passivity assumption suggest that the perception of an animal neonate is like that of a human neonate: meaningless and unorganised; hence, animals—like humans—have to learn to perceive. The empirical evidence against this claim was discovered earlier than in the case of human neonates. (This is so because it is more obvious in animals, and because of the use of experiments which were not available, for moral reasons, in the case of human neonates.) Empirical research shows that some animal neonates perceive distance, form, size constancy, brightness constancy, and other features which are regarded (in the sensation-perception distinction) as part of the later stage of 'contaminated' perception. Hence, 'the distinction between sensation and perception appears to be entirely unwarranted' in the case of animal neonates (Hochberg [1962], pp. 304–5, 322–5; see also Pastore [1971], pp. 203–6, 312–19).

Empirical research clearly failed to find a stage that can be identified as the stage of pure sensation. That is, there is no empirical evidence supporting the passivity assumption. For this reason most of the advocates of the sensation–perception distinction in our century do not conceive of pure sensation as an empirical stage, but rather a theoretical one. Indeed, theoretical considerations have played a crucial role in the postulation of the passivity assumption and its sensation–perception distinction. My criticism of the theoretical postulation of pure sensation is twofold: (a) it is meaningless to speak about such a theoretical stage; (b) this stage does not solve the theoretical difficulties its postulation is intended to.

Even if pure sensation is regarded as a theoretical entity, its positive features should be pointed out. But besides the negative requirement concerning the absence of cognitive, emotional and evaluative contributions of the agent, no positive attribute can be given to pure sensation. As Collingwood [1923] puts it: 'In an act of pure sensation we should not be remembering, imagining or thinking: we should be aware of nothing but that which was directly before us, as immediate object, at this precise moment' (p. 58). But excluding the above, cognitive features prevent us from putting pure sensation in any conceivable context! Consequently, we can say nothing more, 'for to say anything about it, even to ourselves, means placing it in a context of objects which are not present to pure sensation and are therefore, so far as we are purely sensating, not present to our minds at all' (p. 59). We cannot even say that we are confronted with a coloured patch, since to call something 'colour' implies to distinguish it from smell or sound or something else that is not coloured; such a distinction cannot exist in the stage of pure sensation. No positive attribute can be given to the abstract notion of pure sensation. (No wonder the advocates of the notion ascribe it with conflicting attributes.) This notion must be rejected because it has no empirical basis and no meaning whatsoever:

> A real sense-datum would be, if it existed, just nothing at all; it would be unsusceptible of any kind of description, indistinguishable from any other sense—datum or anything not a sense–datum. Can there be a completer description of an absolute nonentity? (Collingwood [1923], p. 62).

In general, it is inconceivable that a system (*viz*. the mind) which is defined by its meaningfulness, has a meaningless stage (*viz*. pure sensation).

So far I have criticised the postulation of pure sensation on the grounds that it is not compatible with empirical evidence or any description at all. In the light of this, the question of whether the theoretical notion of pure sensation can overcome the difficulties in the conceptual framework of the seventeenth century has an obvious negative answer; an empty and non-descriptive notion cannot solve any problem. Nevertheless we shall briefly discuss this question since the stage of pure sensation was originally postulated because of theoretical considerations and not in order to present a more adequate empirical description of the perceptual process.

Even if the notion of pure sensation is not as empty, as I believe it is, it does not solve the difficulties which emerge within the conceptual framework of the seventeenth century. As mentioned above, pure sensation is considered 'the intersection of mind and matter' and as such it is regarded as an intermediate stage in the causal transition in the head from physical events to mental ones. This kind of solution is similar to being only a little bit pregnant. That is, giving a quantitative solution to a qualitative gap. The critical problem remains intact: How can physical and mental entities be on the same causal chain?

The stage of pure sensation does not have epistemic value as well. If it is a meaningless stage devoid of any cognitive meaning it cannot be the starting point of the cognitive process. The cognitive system cannot construct a meaningful world out of completely meaningless materials. In the same way that one cannot construct buildings out of mere ideas, one cannot construct a meaningful world out of meaningless materials. In both cases, the construction process must use relevant raw materials.

The passivity assumption that postulates the stage of pure sensation was absent in ancient philosophy since the conceptual difficulties stemming from the modern mind-body dichotomy by and large were absent as well. The passivity assumption was the first attempt to counter these difficulties; I have shown that it was unsuccessful. However, we cannot go back and adopt the ancient conceptual framework based on naive realism. The difficulties connected with the mind-body dichotomy in general and the explanation of the perceptual process in particular have not disappeared. Hence, we cannot adopt a conceptual framework which, generally speaking, ignores them. We need a novel conceptual framework that would recognise these difficulties and tackle them differently.[1] In particular this framework would have to abandon the attempt to find an intermediate stage in the causal transmission between physical and mental events. This attempt (as expressed in the passivity assumption), though advocated since the seventeenth century, has serious empirical and theoretical flaws.

University of Haifa

---

[1] Suggestions for such framework can be found in Strauss [1984] and Ben-Zeev [1983, 1984a].

REFERENCES

BEN-ZEEV, A. [1983]: 'Toward a Different Approach to Perception', *International Philosophical Quarterly*, **23**, pp. 45–64.

BEN-ZEEV, A. [1984a]: 'The Kantian Revolution in Perception', *Journal for the Theory of Social Behaviour*, **14**, pp. 69–84.

BEN-ZEEV, A. [1984b]: 'Aristotle on Perceptual Truth and Falsity', *Apeiron*.

BEN-ZEEV, A. and STRAUSS, M. [1984]: 'The Dualistic Approach to Perception', *Man and World*, **17**, pp. 3–18.

BERKELEY, G. [1713]: *Three Dialogues between Hylas and Philonous*, in *Philosophical Works*, edited by M. R. Ayer. Dent [1975].

BERKELEY, G. [1733]: *The Theory of Vision Vindicated.*

BLOCK, I. [1960]: 'Aristotle and the Physical Object', *Philosophy and Phenomenological Research*, **21**, pp. 93–101.

BOWER, T. G. R. [1982]: *Development in Infancy*; 2nd ed. Freeman.

CASSIRER, E. [1910]: *Substance and Function*. Dover [1953].

COLLINGWOOD, R. G. [1923]: 'Sensation and Thought', *Proceedings of the Aristotelian Society*, **24**, pp. 55–76.

DESCARTES, R. [1628]: *Rules for the Direction of the Mind*, in *The Philosophical Works of Descartes*, edited by E. S. Haldane and G. R. T. Ross. Cambridge University Press [1955].

DESCARTES, R. [1641]: *Objections and Replies.*

DIDEROT, D. [1965]: *Encyclopedia, Selections*. Bobbs-Merrill.

DRETSKE, F. [1969]: *Seeing and Knowing*. The University of Chicago Press.

HALLECK, R. P. [1895]: *Psychology and Psychic Culture*. American Book.

HATFIELD, G. C. and EPSTEIN, W. [1979]: 'The Sensory Core and the Medieval Foundations of Early Modern Perceptual Theory', *Isis*, **70**, pp. 363–84.

HELMHOLTZ, H. VON [1867]: *Treatise on Physiological Optic III*. The Optical Society of America [1925].

HELMHOLTZ, H. VON [1878]: 'The Facts of Perception', in *Selected Writings of Herman von Helmholtz*, edited by R. Kahl. Wesleyan University Press [1971].

HOCHBERG, J. E. [1962]: 'Nativism and Empiricism in Perception', in *Psychology in the Making*, edited by L. Postman. Knopf.

HOCHBERG, J. E. [1978]: *Perception*; 2nd ed. Prentice-Hall.

KÖHLER, W. [1929]: *Gestalt Psychology*. Mentor [1975].

LEWIS, C. I. [1929]: *Mind and the World Order*. Dover [1956].

LINDBERG, D. C. [1976]: *Theories of Vision from Al-Kindi to Kepler*. The University of Chicago Press.

LOCKE, J. [1690]: *An Essay Concerning Human Understanding*. Dover [1959].

MELZACK, R. [1973]: *The Puzzle of Pain*. Basic Books.

PASTORE, N. [1971]: *Selective History of Theories of Visual Perception: 1650–1950* Oxford University.

PLANCK, M. K. E. L. [1925]: *A Survey of Physics*. Methuen.

PRIESTLEY, J. [1777]: *Disquisitions Relating to Matter and Spirit*, in *Priestley's Writings*, edited by J. A. Passmore. Collier [1965].

RUSSELL, B. [1921]: *The Analysis of Mind*. George Allen and Unwin.

RUSSELL, B. [1927a]: *The Analysis of Matter*. Paul, Trench and Trubner.

RUSSELL, B. [1927b]: *Philosophy*. Norton.

RUSSELL, B. [1940]: *An Inquiry into Meaning and Truth*. George Allen and Unwin.

RUSSELL, B. [1948]: *Human Knowledge*. Simon and Schuster.

RYLE, G. [1949]: *The Concept of Mind*. Hutchinson.

SCHAFFER, L. F., GILLMER, H. VON, and SCHOEN, M. [1940]: *Psychology*. Harper.

SHAW, R. E., TURVEY, M. T. and MACE, W. [1981]: 'Ecological Psychology: The Consequence of a Commitment to Realism', in *Cognition and Symbolic Processes, II*, edited by W. Weimer and D. Palermo. Lawrence Erlbaum.

SMITH, N. W. [1974]: 'The Ancient Background to Greek Psychology and Some Implications for Today', *Psychology Record*, **24**, pp. 309–24.

SPINOZA, B. [1677]: *The Ethics*, in *The Chief Works of Benedict De Spinoza*, edited by R. H. M. Elwes. Dover [1955].

STRAUSS, M. [1984]: *Empfindiung, Intention und Zeichen: Typologie des Sinntragens*. Alber.

TOULMIN, S. [1967]: 'Neuroscience and Human Understanding', in *Neuroscience: A Study Programme*, edited by G. Quarton, F. Schmitt and E. Melnechuk. Rockefeller University Press.

TOULMIN, S. [1981]: 'Evolution, Adaptation, and Human Understanding', in *Knowing and Validating in the Social Sciences: A Tribute to Donald T. Campbell*, edited by M. Brewer and B. Collins. Jossey-Bass.

TOULMIN, S. [in press]: *Mastery and Adaptation* (*Human Understanding: Vol. II*).

WALK, R. D. [1978]: 'Perceptual Learning', in *Handbook of Perception, IX*, edited by E. C. Carterette and M. P. Friedman. Academic Press.

WUNDT, W. [1907]: *Outline of Psychology*. Engelmann.

Brit. J. Phil. Sci. 35 (1984), 345–358 Printed in Great Britain

Individualism and Global Supervenience*

by GREGORY CURRIE

1 *Introduction*
2 *The Individual and the Social*
3 *The Definition*
4 *The Thesis*
5 *A Problem about Belief*
6 *Ontological Individualism*

1 INTRODUCTION

My aim in this paper is to formulate and defend a metaphysical principle which may help us to overcome a misleading dichotomy between individualistic and holistic ways of thinking about the social sciences. The idea is that there is a non-causal, non-reductive relation of dependence between facts about social institutions and facts about the behaviour of individuals; that the social is *supervenient upon* the individual. Adopting this idea enables us to reconcile certain holistic aspects of the social—aspects which I shall discuss below—with an insight that has motivated much that has been written in defence of methodological individualism (*MI*). This insight has been expressed in various ways: that social institutions 'do not possess a quasi-divine autonomy'; that 'every complex social situation or event is a result of a particular configuration of individuals, their dispositions, situations, beliefs and physical resources and environments'; that 'it is *people* who determine history, however people themselves are determined'.[1]

The supervenience thesis captures these ideas in a quite precise way. It says that, whatever complex and reciprocal relations there are between social entities and individuals, it is the totality of individual facts which determines the totality of social facts. Accordingly, I shall begin with a distinction between individual and social facts. I shall then offer a rigorous formulation of the supervenience thesis. It will be shown how the thesis

Received 27 May 1983
Final version received 6 December 1983

---

* I am grateful to a number of people who have commented on various parts of this paper, including Loraine Hawkins, Frank Jackson, Alan Musgrave, Kim Sterelney and, in particular, Graham Oddie. Conversations with David Armstrong helped me to formulate some of the ideas presented here. Criticisms of an earlier version by a referee of this *Journal* enabled me to improve the paper in various ways.

[1] These quotations are from Watkins [1952], p. 149; [1957], p. 168 and [1955], p. 179 respectively. (Italics in the original.)

handles a problem about the social determination of belief. Finally I shall consider the relation of the supervenience thesis to another metaphysical principle often thought to be constitutive of *MI*: that social entities are constructions which have individuals as their parts or members. The rejection of this principle will be seen to be compatible with supervenience. Thus we have the dependence of social facts and the irreducibility of social objects.

Supervenience theses have proved useful in other areas of philosophy. It has been said that, while the various kinds of value (moral, aesthetic) that an object may have are not reducible to its naturalistic properties, they are supervenient on its naturalistic properties in the sense that two entities, not differing in their naturalistic properties, may not differ from the point of view of moral and aesthetic value. And recent failures to give a reductive solution to the mind-body problem have caused some philosophers to retreat to the weaker but still contentious view that the mental is supervenient on the physical; that while it may not be possible to give definitional or nomic equivalents of mental properties in physical terms, the application of these properties is still strongly dependent upon purely physical facts. For the thesis that the mental is supervenient on the physical is the claim that physically indistinguishable beings will also be mentally indistinguishable. The physical is the independent variable and the mental the dependent variable.[1]

There is not, however, common agreement about how to specify the supervenience relation. The formulation I shall offer here may, I think, be useful in these other areas. So while I am using the idea of supervenience to solve a problem about the social sciences, I am also using the example to probe the resources of supervenience.

2 THE INDIVIDUAL AND THE SOCIAL

In order to speak of the supervenience of the social on the individual we must be able, in a preliminary way, to distinguish the two. The distinction I shall make will not be very precise and will depend heavily on examples. But lack of precision is not fatal to us here. As with the mental and the physical, our conviction that there is, at least at the conceptual level, a distinction to be made is stronger than our faith in any particular proposal as to how the distinction is to be drawn. And those who argue about whether the mental is supervenient on the physical seem to make progress without it being absolutely clear how or where they draw the distinction between the two.

Social facts I take, roughly speaking, to be facts about social institutions and roles, and facts about people's actions, where those actions have a social significance. That unemployment rose last year, that the local store changed hands last month, that Smith voted Labour at the last election—all these are social facts. By individual facts I mean facts about the inner mental states (both cognitions and sensations) of individuals and facts about their bodily movements. That Smith wanted to vote, that Smith raised his arm, are individual facts about Smith, while the fact that he voted is a social fact about him. For arm raising brought about by the desire to vote does not constitute voting; we need, in addition, a certain social setting. I hope that the significance of this distinction will emerge more fully as the discussion proceeds.

---

[1] For the supervenience of the moral see Hare [1952], p. 145. The supervenience of the aesthetic is discussed in Walton [1970]. Davidson proposed the supervenience of the mental on the physical in his [1970], p. 88, and [1973], pp. 716–17. See also Kim [1978] and [1979] and Haugerland [1982].

It must be acknowledged that the specification of a person's mental states requires unrestricted reference to social concepts. For the content of many, perhaps most of our beliefs (and other propositional attitudes) involve such concepts. To believe that the banks are closed on Sunday, that the Conservatives are in power, to desire a better job or a new car are all obvious examples of this. Yet facts of this kind I count as individual facts. What is the justification for this?

First of all, when advocates of *MI* spoke of 'the dispositions and beliefs of individuals', in terms of which they hoped to explain social events,[1] they did not suppose that those beliefs could be specified in non-social terms. They did not suppose, in other words, that every social concept has a non-social synonym. In fact they have explicitly denied that they were seeking a reduction of social concepts to non-social ones. Secondly, the attribution to an individual of a thought involving a social concept does not entail any particular social fact, because of the obliquity of expressions in 'that' clauses. That I believe that the Prime Minister is a woman does not entail the existence of a Prime Minister. So the fact that social concepts are essentially required as constituents of propositional attitudes is not a way of smuggling social facts into the domain of the individual.

Tyler Burge has argued that beliefs are essentially social, and I shall consider his argument in section 5 below.

3 THE DEFINITION

Later I shall use the distinction just drawn between individual and social facts in our formulation of the supervenience thesis. But I want to begin with a quite general discussion of supervenience. So let N be a class of facts, events or characteristics, and let M be another such class. Then M supervenes on N (or, N is a supervenience base for M) if and only if sameness with respect to N entails sameness with respect to M. I will now make this idea more precise.

Supervenience I take to be an ontological relation between classes of facts. (A definition in terms of classes of properties would, I think, do equally well.) I shall therefore avoid formulating the thesis in terms of relations between theories or classes of predicates. Once we have a definition in terms of classes of facts we can proceed to a derivative notion of supervenience in syntactic terms. This will not be my concern here. Given the decision to treat supervenience as a metaphysical relation, the idea naturally suggests itself of using the machinery of possible worlds. I take a possible world to be a counterfactual situation, or a way the world might have been. Worlds differ from one another in respect of the facts which hold in them. Two distinct worlds must differ with respect to the facts, but they may agree on some proper subset of them. Much more could be said by way of analysis and defence of possible world semantics, but it is clearly inappropriate to go into such matters here. Also, I have tried to leave my characterisation sufficiently imprecise so as to be consistent with a number of different detailed conceptions of how possible worlds are to be presented. Any one of the currently available explications of that notion could be applied in what follows.

---

[1] See Watkins [1955], p. 179.

Do we really require the machinery of possible worlds? There are two reasons for thinking that we do. First of all, supervenience theses are naturally construed as having a certain modal force, and modalities are best explained in terms of possible worlds. Secondly, in order to accommodate certain holistic features of the social, I want to consider supervenience relations between classes of facts, not as they relate to a particular object in a domain, but as they concern the *whole* domain of objects. In that case we need, instead of a comparison of particular objects with respect to the facts about them, a comparison of domains with respect to the classes of facts which hold of all objects in them. And a possible world is individuated exactly by the facts which hold of a domain of objects. Some of this will become clearer if we consider a formulation of supervenience, due to Kim [1978], which does not involve possible worlds.

We will write '$S(M,N)$' for 'M is supervenient on N' and '$I_N(x,y)$' for 'x and y are N-indiscernible', which means that x and y share the same properties in N. Then we have

$$S(M,N) \equiv_{df} \boxed{?}\, \forall x \forall y (I_N(x,y) \supset I_M(x,y)).$$

I shall make some comments on this definition. First of all, I write '$\boxed{?}$' because Kim leaves the modal status of the supervenience thesis unclear. (He is concerned with the putative supervenience of the mental on the physical.) I shall formulate my own thesis in such a way that the modal operator is strict logical necessity. Secondly, this definition is deficient from my point of view. The right hand side involves the idea that the N- and M-properties belong to the same objects: if two objects agree in their N-properties they will agree in their M-properties. So Kim's formulation confines us to talking about individual and social properties which we attribute to the same objects. This is too restrictive. Social institutions can be treated as objects with social properties, but they cannot be said to have individual properties (*e.g.* they do not have mental states). We want the idea that the *N*-properties of one set of objects determines the *M*-properties of another set of objects (though of course the two sets may overlap.)

Another difficulty arises when we consider what sort of properties might belong to the *M*- and *N*-classes in Kim's definition. (This brings me to an explanation of the holism mentioned three paragraphs back.) When we say, for instance, that *a* and *b* share the same physical properties, what is to count as a physical property? Clearly we want to include intrinsic physical properties like molecular structure, but we might also want to include broadly relational, or extrinsic, physical properties, like being 200 miles from the centre of mass of the solar system at time *t*. These kinds of extrinsic properties can be important in formulating supervenience theses. For example, some aestheticians want to argue that even if two pictures were alike in their intrinsic physical properties (so that there would be no physical tests we could carry out to determine which is which) they may still not be alike in their aesthetic properties. The reason is that the aesthetic properties of the work depend partly on its history, and two apparently identical paintings can have different histories (see Walton [1970]). In order to include the history of the work we would have to build in broadly relational properties when we consider the requisite notion of sameness. Similarly, when we turn to the alleged supervenience of the mental on the physical, there is reason to think that the thesis is plausible only if we include, in the condition of physical sameness, relational properties. For Putnam [1975] has argued that a person's psychological state is not a matter simply of what is going on 'in his head' but also of what external objects he is causally related to.

The same difficulty arises when we consider facts about the social. There is an undeniable sense in which such facts are holistically constituted. My becoming Prime Minister is not just a matter of what I think and do; it depends upon what others think and do as well. So my social characteristics are clearly not determined by my individual characteristics alone. If individual facts determine the social facts, they do so in a global rather than a local way. In other words, we need to formulate the thesis somewhat like this:

(1) The totality of social facts is supervenient upon the totality of individual facts.

In order to make this more precise, we will introduce some further concepts. In each possible world we distinguish two classes of facts; individual and social facts. I shall make the desired distinction within a world *w* by talking about the individual facts in *w* and the social facts in *w*. The totality of individual facts in *w* will be called 'the individual history of *w*'. Similarly for the totality of social facts in *w*. We can now give a preliminary formulation of the claim that the social is supervenient on the individual:

(2) $\forall u \forall w$ (if u and w have the same individual history then they have the same social history).[1]

(2) can be strengthened by bringing time into the picture. At any time t we can 'freeze' a world w and ask what individual and social facts hold in w at t. Let us call the individual facts which hold in w at t 'the individual state of w at t'. Similarly for social facts. Now worlds which have the same individual or social state at t for some t may yet differ in their individual or social histories; they coincide momentarily, so to speak. Should we say that worlds with the same individual state at t have the same social state at t? This is too strong. What determines the social state of w at t is not simply the individual state of w at t. The social significance of the actions of individuals in w at t may depend on the prior establishment of conventions by those and other individuals. So perhaps we should weaken our thesis to say that if two worlds are such that they have the same individual histories up until t, then their social states at t will be the same. The following is then what it means to say that the social supervenes on the individual:

(3) $\forall u \forall w \forall t$ (if u and w have the same individual histories up to t then they have the same social states at t).

Supervenience in the sense of (3) gives a rather minimal interpretation to the claim that the social depends on the individual. Should we look for something stronger—bearing in mind, of course, the holistic constraints which lead to the rejection of Kim's formulation? In fact Terence Horgan, addressing exactly this problem of holism, has proposed a somewhat stronger interpretation (see Horgan [1982]). I shall briefly outline Horgan's suggestion and give a reason for thinking that it is inadequate.

Horgan's method is to use the idea that facts or events are intrinsic to certain spatio-temporal regions and not to others. My remembering Vienna is intrinsic to the area occupied by the earth during my lifetime (and to some smaller regions as well) but not intrinsic to the region occupied exclusively by my body over the same period of time. My remembering Vienna depends logically on my having been there. He says that 'a feature is intrinsic to a given region if its presence does not depend, in a broadly logical sense of "depend", upon what happens outside the region' (*ibid.*, p. 36). Let M and N be classes of facts. Then, neglecting certain features of Horgan's discussion irrelevant to our topic, Horgan's formulation of the claim that M supervenes on N is this:

(4) There do not exist any two regions exactly alike in all intrinsic N-facts but different in some intrinsic M-fact.

---

[1] A similar principle is proposed in Hellman and Thompson [1975], p. 558. Their formulation is model-theoretic and involves reference to a language. Macdonald and Pettit, in their [1981], assert the supervenience of the social on the individual without offering a precise explication of the notion. They base the thesis on the idea that social wholes have individuals as their parts. I reject this idea. (See *below*, section 6.)

In order to understand this proposal let us apply it to the case of the mental and the physical. Suppose that Fred is thinking of Vienna, and his doing so is intrinsic to the region *r*. Let *P* be the class of physical facts intrinsic to *r*, and let *r*′ be another region to which qualitatively the same physical facts are intrinsic. Then, by Horgan's formulation of the supervenience thesis, the individual in that region *r*′ who is the counterpart of Fred will also be thinking of Vienna.

The first difficulty is this. Horgan says that a 'feature' is intrinsic to a region if it does not logically depend upon what happens outside that region. A 'feature' of a region is, presumably a fact which holds, or an event which occurs, within that region, so Horgan seems to be telling us that an event is intrinsic to a region if it depends only on events intrinsic to that region, and this is clearly circular. If we follow Kim in taking an event to be the instantiation of a property (relation) by an individual (*n*-tuple of individuals) at a time, then the following way out is suggested. Let us say that event *a* logically depends on event *b* if the occurrence of *a* entails the occurrence of *b*. In this sense my remembering Vienna logically depends on my having visited that city. If *e* is an event let *D*(*e*) be the set of events upon which *e* logically depends (*e* depends upon itself), and let *O*(*D*(*e*)) be the class of objects constitutive of the events in *D*(*e*). Then we can say that *e* is intrinsic to a region *r* if the objects in *O*(*D*(*e*)) are all contained within *r*. But now we come upon another difficulty. If we regard cognitive attitudes like beliefs and desires as relations between the thinking subject and a proposition, and if we treat propositions as abstract entities without spatio-temporal location, then it will be impossible, on this definition, to locate cognitive acts within spatio-temporal regions. But facts about people's cognitions are paradigmatic examples of individual facts. If we cannot localise them, we cannot hope to use the idea of intrinsicality in our definition of supervenience.

The claim that cognitive attitudes are attitudes towards abstract propositions is certainly not universally accepted. The attempt is often made to replace them by sentences 'in the language of thought'. I do not think that such attempts can succeed, but it must be made clear that I am rejecting Horgan's proposal (as modified) from a certain philosophical perspective. Perhaps the analysis of cognitive attitudes adopted here is wrong and Horgan's proposal, or something like it, can be accepted. I think that the remaining results of this essay could, with non-essential modifications, be translated into his framework. I shall not attempt this here.

4 THE THESIS

Our thesis is, then, that the social is supervenient on the individual in the sense of (3). Resistance to such a thesis on the part of holists should, I believe, dissipate now that its thoroughly global character is made clear. If we concentrate on the relation of an individual or relatively small group of individuals to an institution we are inclined to assert the autonomy of the institution's states; to assert that there is reciprocal action and no order of priority between individuals and the institution. What happens to a bank is not wholly determined by the actions of its directors, nor even by the actions of all its employees, owners and clients. It bears relations to other social institutions which affect its states. This is, of course, perfectly compatible with global supervenience, which demands merely that, if there is to be a social difference between worlds *u* and *w* (at time *t*) there must be some difference between them concerning the thought or behaviour of at least some individual (up to *t*). A difference between *u* and *w* not reflected in any individual differences would simply not count as a social difference. While, as I shall argue, social concepts are irreducible to individualistic concepts, their application is highly sensitive to the application of individualistic ones. For there simply would be no content to the claim that while worlds *u* and *w* are absolutely indistinguishable with respect to what individuals are thinking, perceiving, saying and doing, there is a difference between them in terms of how some social concept is to be applied.

Now it might be objected that the supervenience of the social on the individual does not establish an order of metaphysical priority of the one over the other, if it turns out that we can invert the relationship and show that the individual is supervenient on the social. Can this be done? Must worlds which are socially indistinguishable be individually indistinguishable? No. It is possible for worlds which have the same social histories up to *t* to be in distinct individual states at *t*. The reason is that social characteristics display what has been called (in another context) 'compositional and configurational plasticity'.[1] That is, many different combinations of individual thoughts and actions may realise the same social state. The beings who are members of society might be very different from human beings (for instance in their physical constitution) without that difference affecting the social facts. Just as the same mental state can be realised by a Martian and by a human, so the same state of society can be realised by a community of Martians and by a community of human beings. Nor is it the case that there need be a psychological isomorphism between the two communities for them to realise the same social state. Two reasons for this are worth mentioning. First of all, action within an institutional context tends to be characterised largely by its behavioural content. A person will count as having voted for the motion if and only if he raises his arm at a certain time (given a convention connecting arm-raising with voting), independently of the intention with which he raised his arm. He might mistakenly have thought that he was voting against the motion, or have been intending to ask a question. Thus, for the motion to be carried, it is not necessary for a majority of voters to have intended to vote in favour of it. Secondly, the relation between mental states and actions is itself holistically determined.

---

[1] See Boyd [1980], pp. 87–8.

Beliefs and desires give rise to actions only as they are modified and enriched by other beliefs and desires (see Davidson [1970]). So, even barring misunderstandings such as described above, those voting in favour of the motion may not all have the same guiding beliefs and intentions. A person may want the motion to fail, but feel an obligation to vote for it, or hope that in so voting he will encourage others to vote against it.

It might be argued that such individual differences, while they may make no difference to the outcome of a particular social process, will make a difference, if only a minor one, to the outcome of some other social process—we are, after all, considering whether the individual is *globally* supervenient on the social. But there is no reason to think that differences in individual attitude will always show up somewhere at the social level. For attitudes are socially relevant only if they find behavioural expression (*e.g.* linguistic expression). A belief that an individual has is not socially relevant unless other members of the community have some reason to attribute that belief to the individual. A belief which never finds any kind of expression, no matter how indirect, remains part of that individual's subjective world. And since the relation of mental states to behaviour is a contingent one, it is possible for a particular mental state never to show up at the level of behaviour (see Putnam [1965]). Further, certain differences at the level of bodily movement will be socially irrelevant (consider slight differences in arm movements at voting time). So worlds may differ at the individual level both in respect of thought and behaviour without differing socially.

The plasticity of social concepts, as well as guaranteeing that the individual is not supervenient on the social, guarantees that there cannot be any methodological reduction of the social to the individual. Social predicates like 'is a bank', 'is a cohesive society', *etc.*, cannot be replaced by (finite conjunctions and disjunctions of) predicates of individuals. For the content of a social predicate is never exhausted by any given combination of individual actions. It is always possible for some other such combination to realise the same state of an institution.[1] But while supervenience enables us to eschew any reductive relation between predicates of the language in which we express social and individual theory, it does involve certain relations of entailment, if we make a further appeal to possible world semantics. We identify propositions with classes of worlds: the proposition P is the class of worlds in which P is true.[2] Now consider a world w and the class of worlds W which have the same individual history as w. Let P be the proposition true in exactly all members of W. P expresses the individual facts about w. Let W' be the class of worlds with the same social history as w, and let Q be the proposition true in exactly all the members of W'. Q expresses the social facts about w. Now, by the supervenience thesis, P

---

[1] This argument goes back to Putnam [1967]. See also Papineau [1985].

[2] This gives us an analysis of propositions too coarse for the purposes of individuating propositional attitudes, but fine enough for dealing with entailment relations.

entails *Q*. For a proposition *P* entails a proposition *Q* if every world in which *P* is true is a world in which *Q* is true. And the supervenience thesis says that *Q* is true in every world in which *P* is true.

This result is of little methodological interest: for worlds of any complexity the proposition which expresses the individual history of that world is unlikely to be one which we could ever know, or even express in a human language. Nor can we use the result to make the conditional epistemic claim that if we knew all the individual facts about the world we would know all the social facts about the world. For knowledge is not closed under entailment.[1]

5 A PROBLEM ABOUT BELIEF

Our global formulation of the supervenience thesis allows us to cope with a problem about the social determination of belief which has recently been discussed, and which would prove a difficulty for stronger versions of the thesis. Tyler Burge [1979] has argued that the ascription of propositional attitudes is dependent upon the subject's social environment. More specifically, two subjects might have the same qualitative experiences, physiological states and dispositions to behave, and be responding to the very same stimuli, but have distinct attitude contents (see *ibid.*, p. 79). Briefly, Burge's argument is that what terms it is correct to employ in characterising the subject's attitudes depends upon the rules prevailing in the community for the correct application of the term. And differences between communities with respect to such rules may not be reflected in any of the aforementioned features of the subject and the perceptual-causal relations he bears to the environment. Thus it may be that Jones_1, and his counterpart Jones_2 have the same history of mental and bodily states and relations to their environments, while it is true of Jones_1 that he believes that he has arthritis, but not true of Jones_2 that he believes this; the reason being that there are differences in the (technical) use to which the term 'arthritis' is put in their communities.

This seems to be a very strong argument against individualism because it seems to show that the subject's mental states depend upon social facts about the conventions and knowledge available within a community, even where those facts make no direct causal or perceptual impact on the subject. So one cannot avoid the problem by saying 'it is not the social facts which matter here, but their reflection in the mind of the subject which matter'. But whatever the effect of Burge's thesis on some versions of individualism, it evidently presents no counterexample to my global supervenience thesis. For the differences between the situations of Jones_1 and Jones_2 which lead to disparate attributions of propositional attitudes to them must be reflected in differences between the mental states of some other member(s) of the communities in which Jones_1 and Jones_2 live. To see this we should ask what sort of difference it is between the situations of Jones_1 and Jones_2 which make it true of Jones_1 that he believes that he has arthritis and not true of Jones_2 that he believes the same. The difference, according to Burge, is that the medical community in Jones_1's society (call it *M*1) use 'arthritis' for a disease of the joints only, while the medical community in Jones_2's society (call it *M*2) use the same term for a disease which also infects the bones. According to Burge this means that 'arthritis' as used in *M*2 does not refer to arthritis and that Jones_2, when he says 'I think I now have arthritis in my thigh' is not expressing a belief about arthritis. Now what sort of difference is it between *M*1 and *M*2 that justifies us in saying that these communities use 'arthritis' to refer to different things? It must be something manifest in their communicative practice—otherwise it would not be a difference in respect of *use*. Now a difference in communicative practice must be manifested in some difference between the two communities in respect of what is said or written, or in respect of what is communicated by non-verbal behaviour. So the difference must, ultimately, be at the level of bodily movements. We have specified that the bodily movements of a person are individual facts about that person. So we may conclude that Jones_1 and Jones_2, in the situation described, cannot have different beliefs about their medical conditions unless there are individual differences between their communities. And so we see that any differences in their belief states cannot be due to purely social differences between their communities. The demand for *global* supervenience makes the construction of a Burge-type counterexample to our thesis impossible.[1]

---

[1] *E.g.* we do not know all the arithmetical truths which are consequences of the Peano Axioms.

6 ONTOLOGICAL INDIVIDUALISM

The supervenience thesis might serve as the metaphysical 'hard core' of a research programme for the social sciences. Such a programme would recognise the holistic nature of social concepts and their resistance to characterisation in terms of individualistic equivalents (nomic or definitional). But it would affirm the claim that the domain of individual facts is more fundamental than the domain of social facts. Clearly this programme requires articulation in terms of a set of research strategies which cohere with its metaphysics, but that is something I hope to tackle on another occasion. Instead I will ask whether the supervenience thesis ought to be supplemented by a further metaphysical principle.

Recent thinking about physicalism has stressed two principles; the supervenience of the non-physical on the physical and the doctrine that everything is either directly a physical object, or consists wholly of parts which are physical or belongs to the set theoretic closure of the physical. Hellman and Thompson, in their influential study of physicalism without reduction, call this 'the principle of physical exhaustion'.[1] Should we supplement our supervenience thesis with an analogous principle of individual exhaustion (*PIE*): that social entities like institutions are either mereologically or set-theoretically constituted out of individuals, and perhaps purely physical objects as well? One idea might seem to force us to that very conclusion: the Humean principle that 'distinct existences' must be logically independent.[2] If the states of social entities are supervenient in our sense on the states of individuals then their states are logically determined by the behaviour of individuals. But then, by the principle of independence (as I shall call Hume's idea), social entities must be identical with, or constructions out of, individuals.

---

[1] A referee's comment enabled me to improve this part of the argument.

Is this conclusion really forced on us? I believe not. Accepting the supervenience thesis and a version of the principle of independence, I reject *PIE*. I shall not, however, argue that *PIE* is false, because David-Hillel Ruben has recently made a persuasive case to that effect. He argues that a social entity like France cannot be identified with any particular geographical region, with any set or group of individuals, or with a sum whose parts are individuals.[3] This effectively rules out all the strategies within the purview of *PIE*.

Sound though I think his arguments are, I shall not rely upon them here. For I construe the question hypothetically. *If PIE* is false, must we abandon the conjunction of the supervenience thesis and the principle of independence?

First of all we need to give a somewhat sharper meaning to some of our terms. Let us say that x and y are distinct existences if they are not identical, and the one is not a part or member of the other. And we will say that x and y are capable of independent variation if there are worlds u and w such that u and w are indiscernible with respect to the state of x but discernible with respect to the state of y, and worlds u' and w', indiscernible with respect to the state of y but discernible with respect to the state of x. The definitions are, I think, faithful to the original meanings of our terms. Given the first of these definitions we deny *PIE* by asserting

(5) If x is an individual and y a social institution then x and y are distinct existences.

---

[1] Hellman and Thompson [1975]. See also Horgan [1982], p. 36.

[2] Discussions with David Armstrong made me aware of the importance of this principle.

[3] See Ruben [1982] and [1983]. Only at one point is there a weakness in Ruben's argument. Against the view that what is constitutive of France is a set of shared attitudes, he replies that such attitudes will include ones about France. And these will, he claims, presuppose a relation between the believer and France ([1982], p. 306). Clearly this will not weigh with anyone who believes that singular terms in belief contexts are non-referential. A stronger argument against the proposal can be had by noting that it would make it impossible for us truly to say—as we surely can—such things as 'If the French Revolution had not occurred, French attitudes towards democracy would be very different.' Nor is it clear how the proposal could handle statements like, 'The geographical area occupied by France is greater than that occupied by Belgium', since either a shared set of attitudes does not occupy a geographic region or, if it can be said to occupy a region in some extended sense, it is unlikely to occupy exactly the region occupied by France.

The following is a plausible version of the principle of independence:

(6) If x and y are distinct existences then the states of x and y are capable of independent variation.

Finally we have

(7) The social is supervenient on the individual.

I claim that (5), (6) and (7) are jointly consistent. Certainly, from (5) and (6) it follows that

(8) If x is an individual and y a social institution the states of x and y are capable of independent variation,

and this may be thought to be inconsistent with (7), which says, surely, that the states of institutions are *not* capable of variation independently of the states of individuals. But of course it is *global* supervenience that is being asserted in (7), not that the behaviour of a single individual ever determines the states of a particular institution, which is what (8) denies. Consider worlds u and w which are indiscernible with respect to the state of individual x. They need not, according to (7), be indiscernible with respect to the states of institution y, as long as there are differences between them expressible in terms of the behaviour of individuals other than x. I conclude that we are free to hold that the social is (globally) supervenient on the individual, to assert Hume's principle of independence and to believe that social institutions are ontologically quite distinct entities from collections of individuals.

If social institutions are, as I believe, ontologically independent entities it seems that clear-cut identity conditions must be provided for them. Institutions are individuated, I suggest, by the effects they have on individuals. A bank affects its owners, directors, employees and clients in various ways. It has effects of a more distant kind on individuals less closely related to it. (Sometimes these less direct effects can be quite dramatic, as when the bank fails.) It would be plainly absurd to insist that, while institutions a and b were indistinguishable with respect to their effects on the totality of individuals, they were yet *distinct* institutions. One can, of course, imagine distinct institutions with remarkably similar effects. Two banks might, conceivably, have the same directors, employees, shareholders and clients. But if we further insist that their effects be precisely identical; that every occasion on which an employee worked for a was an occasion on which he worked for b, that every withdrawal from a was a withdrawal from b, *etc.*, then surely there would be nothing which distinguished the one from the other. This idea of absolute sameness of effects could be made more precise, but it would take us beyond the concerns of this essay to do so here.

In sum then, there is a global relation of supervenience between the individual and the social, and this relation is compatible with the idea that reference to social institutions is irreducible to reference to classes or collections of individuals.

University of Otago

REFERENCES

BOYD, R. [1980]: 'Materialism without Reductionism', in N. Block (*ed.*), *Readings in Philosophical Psychology*. Harvard University Press.

BURGE, T. [1979]: 'Individualism and the Mental', in P. French *et al.* (*eds.*), *Midwest Studies in Philosophy*, 4, pp. 73–121.

DAVIDSON, D. [1970]: 'Mental Events', in L. Foster and J. W. Swanson (*eds.*): *Experience and Theory*. University of Massachusetts Press.

DAVIDSON, D. [1973]: 'The Material Mind', in P. Suppes *et al.* (*eds.*), *Logic, Methodology and the Philosophy of Science*. North Holland.

HARE, R. [1952]: *The Language of Morals*. Oxford: Clarendon Press.

HAUGERLAND, J. [1982]: 'Weak Supervenience', *American Philosophical Quarterly*, **19**, pp. 93–103.

HELLMAN, G. and THOMPSON, F. [1975]: 'Physicalism: Ontology, Determination and Reduction', *Journal of Philosophy*, **72**, pp. 551–64.

HORGAN, T. [1982]: 'Supervenience and Microphysics', *Pacific Philosophical Quarterly*, **63**, pp. 29–43.

KIM, J. [1978]: 'Supervenience and Nomological Incommensurables', *American Philosophical Quarterly*, **15**, pp. 149–56.

KIM, J. [1979]: 'Causality, Identity and Supervenience in the Mind-Body Problem', *Midwest Studies in Philosophy*, **4**, pp. 31–49.

MACDONALD, G. and PETTIT, P. [1981]: *Semantics and Social Science*. Routledge and Kegan Paul.

PAPINEAU, D. [1985]: 'Social Facts and Psychological Facts', in G. Currie and A. E. Musgrave (*eds.*), *Popper and the Human Sciences*. Martinus Nijhoff. *Forthcoming*.

PUTNAM, H. [1965]: 'Brains and Behaviour', in R. J. Butler (*ed.*), *Analytical Philosophy*, vol. 2. Basil Blackwell.

PUTNAM, H. [1967]: 'Psychological Predicates', in W. H. Capitan and D. D. Merrill (*eds.*), *Art Mind and Religion*. University of Pittsburgh Press.

PUTNAM, H. [1975]: 'The Meaning of "Meaning" ', in *Philosophical Papers*, vol. 2. Cambridge University Press.

RUBEN, D-H. [1982]: 'The Existence of Social Entities', *Philosophical Quarterly*, **32**, pp. 295–310.

RUBEN, D-H. [1983]: 'Social Wholes and Parts', *Mind*, **92**, pp. 219–38.

WALTON, K. [1970]: 'Categories of Art', *Philosophical Review*, **79**, pp. 334–69.

WATKINS, J. W. N. [1952]: 'Ideal Types and Historical Explanation', *British Journal for the Philosophy of Science*, **3**, pp. 22–43. Page references are to the reprint in J. O'Neill (*ed.*), *Modes of Individualism and Collectivism*. Heinemann, 1973.

WATKINS, J. W. N. [1955]: 'Methodological Individualism: a Reply', *Philosophy of Science*, **22**, pp. 58–62. Page references are to the reprint in J. O'Neill (*ed.*), *ibid.*

WATKINS, J. W. N. [1957]: 'Historical Explanation in the Social Sciences', *British Journal for the Philosophy of Science*, **8**, pp. 104–17. Page references are to the reprint in J. O'Neill (*ed.*), *ibid.*

Brit. J. Phil. Sci. 35 (1984), 359–374 *Printed in Great Britain*

A Reasonable Reply to Hume's Scepticism

by RICHARD H. SCHLAGEL

According to Hume, no knowledge attainable by human beings would *ever* justify rational belief in recurrent physical properties and causal effects. He arrived at this conclusion because he denied the possibility of knowing—but not the reality of—either the 'inner natures' or the 'secret powers' of objects which would enable one to intuit or to demonstrate a 'necessary connection' between the internal structures of objects and their observable properties, or between the causal powers of entities and their effects. The purpose of this article is to show that while Hume's scepticism was justifiable in his day given the generally accepted limitations then of scientific inquiry and explanations, it no longer is reasonable in light of later scientific discoveries and theories. To a degree to which no eighteenth-century philosopher or scientist could anticipate, experimental inquiry has disclosed the inherent natures of objects and the underlying causes of phenomena that attest to the interconnectedness of physical reality and thus justify, at least to some extent, our instinctive beliefs in the uniformity of nature and predictability of events.

1 *Introduction*
2 *Hume's Sceptical Position*
3 *His Denial of Necessary Connections in Nature*
4 *His Conception of Causal Necessity*
5 *Critique of Hume's Position*
6 *Historical Basis of His Scepticism*
7 *The Present Unreasonableness of this Scepticism*
8 *Conclusion*

1 INTRODUCTION

Unlike almost all other philosophical notions whose grip on the minds of philosophers has receded with time, Hume's analysis of causation and induction are still commonly regarded as being as valid today as when he first proposed them more than two centuries ago. Indeed, Hume's critical analysis and proposed solution are often cited as the one example of a philosophical contribution that has escaped the usual demolition from critical analysis. Philosophers as diverse as Kant, Strawson, Popper, and Quine each has paid tribute in his own way to Hume's position.

Kant [1783], for example, states that 'Hume . . . demonstrated irrefutably that it was perfectly impossible for reason to think *a priori* and by means of concepts . . . a combination [of cause and effect], for it implies necessity'.

Received 28 November 1983

Strawson [1958] writes: '"If . . . there is a problem of induction, and . . . Hume posed it, it must be added that he solved it. . . ."' Popper [1962], as is his wont, claiming to have 'solved' Hume's problem of induction, asserts: 'There is neither a psychological nor a logical induction. *Only the falsity of the theory can be inferred from empirical evidence, and this inference is a purely deductive one.*' But the greatest compliment has been paid by Quine [1969]: 'On the doctrinal side, I do not see that we are farther along today than where Hume left us. The Humean predicament is the human predicament.'

Given this portentous authority, the reader undoubtedly will regard this article with an understandable degree of Humean scepticism. Nonetheless, I think it can be shown quite convincingly that Hume's critique of causality and induction, like his conception of the origin and structure of knowledge, on which it so fatally depends, though then a stunning insight, *betrays the limited understanding at the time of knowledge in general and of scientific knowledge in particular.* That is, when examined within the context of the explanatory limitations of eighteenth-century science, not only does the rationale for Hume's position become eminently clear, so does the resolution of the problem. Just as his basic analysis of knowledge into atomic impressions and their replication as faint copies or ideas ought to appear to anyone today as simplistic, *so his notion that scientific inferences are limited to the succession of ideas or even to the observable qualities of objects should strike one as definitely contrary to present scientific methodology.* It was the ahistorical or anti-historical approach of recent analytic philosophers, especially the positivists, that prevented them from seeing the solution to Hume's scepticism. If, as Reichenbach claimed, the 'justification' of a view is to be kept completely separate from the context of its 'development', then one is prevented from examining the historically derived presuppositions that contributed to the generation of the problem, presuppositions that often appear less plausible from a later perspective. Yet the realisation that an understanding of the historical context within which a traditional problem was conceived can lead to a clarification of the problem, along with its solution, itself required a revolutionary change of view, a change generally accepted now by philosophers of science.

2 HUME'S SCEPTICAL POSITION

Turning to Hume's sceptical arguments, he was aware that no experience or knowledge of the world acquired in the past would have the least significance for the future but for two beliefs: (1) that what we shall encounter in the future will resemble what we have experienced in the past, and (2) that whatever 'mechanism' or causes produced these similarities will remain the same. Hume never questions the first belief, presumably because it can be easily confirmed, directing his argument against the second. That is, even when we can identify objects and events as being similar to those experienced in the past, we have no *rational* grounds for believing that whatever mechanisms or causes produced those objects and events either remain the same or continue to be operative. As our experience is limited solely to the resemblances and successive appearances of objects—or more accurately to impressions or ideas—we do not experience their underlying causes, therefore without some knowledge of a *necessary* connection with those causes the similarities themselves are not evidence of any continuation or identity in what produces them. As he states in the *Enquiries*:

> In vain do you pretend to have learned the nature of bodies from your past experience. Their secret nature, and consequently all their effects and influence, may change, without any change in their sensible qualities. This happens sometimes, and with regard to some objects: Why may it not happen always, and with regard to all objects? What logic, what process of argument secures you against this presupposition? ([1777], p. 38).

Hume's position implies a strange kind of impressionistic ontology according to which we do not confront any actual causes or productive processes in nature, but merely the surface succession of qualities. All that is productive or causal is hidden: volcanoes, hurricanes, forest fires, tidal waves, or epidemics are not symptoms of discoverable, underlying, dynamic processes and causes, but cinematic sequences the mere succession of which creates the illusion of causal production. Naturally, if we had no evidence *whatsoever* of the dependence of the sensible qualities of objects and their effects on the object's inherent nature, then of course the qualities and effects could be considered entirely disconnected from any inherent properties of the object, such that the former could remain unchanged while the latter changed. But were this the case, one wonders why we ever could have been deluded into thinking that there was a dependent connection in the first place!

To be fair to Hume, perhaps there is a sense in which his view could be considered to be partially correct in that human beings, or, as he would say, 'human nature', tends to confuse familiarity with understanding. When we do not understand a written sentence, for example, we sometimes read it over several times until the mere repetition itself conveys a feeling of comprehension. Similarly, people do tend to delude themselves into believing that they understand the familiar processes of nature. As he says:

> The generality of mankind never find any difficulty in accounting for the more common and familiar operations of nature—such as the descent of heavy bodies, the growth of plants, the generation of animals, or the nourishment of bodies by food: But suppose that, in all these cases, they perceive the very force or energy of the cause, by which it is connected with its effect, and is for ever infallible in its operation (*ibid.*, p. 69).

Granting this, there still is a radical difference between the insight that people deceive themselves in thinking that they understand familiar phenomena and the claim that they are deluded in believing that the ordinary occurrences in the world are reliable signs or manifestations of productive causes—in contrast to disconnected, free floating, discrete qualities.

3 HIS DENIAL OF NECESSARY CONNECTIONS IN NATURE

To justify the belief that the experienced similarities in nature are reliable indicators of continuing productive causes we would have to have, according to Hume, knowledge of a 'necessary connection' among such objects or causes and effects as are contiguous in space, successive in time, and constantly conjoined. These latter three relations, while necessary *conditions* for there being necessary connections among objects, are not sufficient to establish such connections. Because rationalists like Descartes [1642] had assumed as the foundation of their rational reconstruction of the world that 'there must . . . be as much reality in the . . . cause as in the effect', Hume is concerned to deny such conceit: 'The mind can never possibly find the effect in the supposed cause, by the most accurate scrutiny and examination' (*ibid.*, p. 29).

Not only is it impossible to find the effect in the cause, we cannot establish any binding tie between the cause and the effect because to do so we would have to be able to 'intuit' or 'demonstrate' an 'invariable' connection between the two. Analogously, to justify the belief that objects resembling those in the past will react in the same manner in the future we would have to obtain either an intuitive or demonstrative knowledge of the inner nature or powers of the object that determines it to behave as it does—a knowledge completely beyond us.

> No object ever discovers [*i.e.*, discloses], by the qualities which appear to the senses, either the causes which produced it, or the effects which will arise from it; nor can our reason, unassisted by experience, ever draw any inference concerning real existence and matter of fact (*ibid.*, p. 27, brackets added).

In addition to the fact that no causal powers are revealed in the qualities of objects, we cannot *infer* such powers from those qualities: 'Should it be said . . . we *infer* a connection between the sensible qualities and the secret powers; this, I confess, seems the same difficulty. . . . The question still recurs, on what process of argument this *inference* is founded (*ibid.*, pp. 36–7)?' It must be emphasized, because many philosophers have misread this, that Hume *is not* denying the *existence* of such secret natures and causal powers, but that we have any basis for claiming *knowledge* of them.

Along with the lack of observational and inferential evidence to support our belief in the necessary connection among causes and effects, Hume relies on another criterion at least as important, that of contradiction. If there were 'necessary connections' evident between causes and effects or between the secret natures or powers of objects and their manifest properties, then it would be contradictory to deny such connections. As he says in the *Treatise*: 'Such a connection wou'd amount to a demonstration, and wou'd imply the absolute impossibility for the one object not to follow, or to be conceiv'd not to follow upon the other: Which kind of connection has already been rejected in all cases' ([1739], pp. 161–2); or, as he says in another passage, 'Such an inference wou'd amount to knowledge, and wou'd imply the absolute contradiction and impossibility of conceiving any thing different' (*ibid.*, p. 87). But as we can always deny the truth or conceive the contrary of any matter of fact or empirical judgment, no necessary connections can be attributed to causal relations or empirical objects: *i.e.*, we can always deny without apparent contradition that the sun will rise tomorrow, that fire will burn paper, or that bread will nourish the human body.

Again, there is an obvious sense in which Hume's view is correct and a deeper sense in which it is incorrect. Obviously, the semantic relation between the terms 'the sun' and 'will rise tomorrow' or 'fire' and 'will burn paper' or 'bread' and 'will nourish the human body' is not as tight as that between 'red' and 'being coloured' or between 'rose' and 'being a plant'. While a denial of the latter expressions clearly would be contradictory, this is not as apparent regarding the former. Yet, as I shall argue more fully later, within the context of Newton's universal law of gravitation and laws of dynamics, *it would be contradictory* to maintain that under the usual physical conditions and application of Newton's laws the earth could cease to rotate on its axis and the sun fail to maintain its orbital position, thereby preventing the apparent rising of the sun.

Similarly, given what we know about the radiant energy of fire and the composition of paper or of the nutritional properties of bread and the digestive processes of a human body, *it would be contradictory* to deny that fire will burn paper and that bread will nourish the human organism. Although the semantic relations between these empirical concepts are not analytical in the usual sense, given what we have discovered about the physical properties of solar bodies (*e.g.*, mass, inertia, gravity, *etc.*) and the nature of fire and paper or bread and the human digestive tract, *it would be inconsistent* to maintain that certain predictable effects would not take place under conditions similar to those that prevailed when the discovery and explanations of these effects occurred. *While not analytically connected, given the experimentally discovered interrelatedness of the empirical concepts of scientific theories, the necessity of certain effects can be inferred, such that their denial would be inconsistent.* As Quine has stated, the strands linking our theoretical network and web of beliefs are not all of the same closeness or tautness, yet they do constitute some degree of semantic connectedness.

According to Hume, for there to be predictable connections among matters of fact they would have to exhibit the same kind of necessity as occurs in mathematics and logic: *i.e.*, the necessity must be self-evident or deductive. Having distinguished sharply the two forms of knowledge, 'relations of ideas' and 'matter of fact', and maintained that the former consists of 'intuitive' or 'demonstrative relations' among certain ideas that can be established by reason independently of experience, Hume surprisingly demands that the network of empirical, scientific ideas display the same kind and degree of analytical or implicative connectedness; then, not finding such necessary connections among our impressions or ideas of observable objects or in matter of fact knowledge, he concludes that the necessity must be subjective, derivable from the mind itself.

> Thus as the necessity, which makes two times two equal to four, or three angles of a triangle equal to two right ones, lies only in the act of the understanding, by which we consider and compare these ideas; in like manner the necessity or power, which unites causes and effects, lies in the determination of the mind to pass from the one to the other. . . . 'Tis here that the real power of causes is plac'd, along with their connection and necessity (*ibid.*, p. 166).

4 HIS CONCEPTION OF CAUSAL NECESSITY

Hume's negative critique of objective necessity leads directly to his positive thesis as to how we subjectively acquire the idea of causal necessity. Given the repetitive similarities among objects and events, human nature is such that our minds habitually or customarily expect these resembling instances to recur. It is this subjective propensity based on recurrent similarities that constitutes the feeling or *impression* of necessity from which the *idea* of necessity is derived.

> The idea of necessity arises from some impression. There is no impression convey'd by our senses, which can give rise to that idea. It must, therefore, be deriv'd from some internal impression, or impression of reflection. There is no internal impression . . . but that propensity, which custom produces, to pass from an object to the idea of its usual attendant. This therefore is the essence of necessity. Upon the whole, necessity is something, that exists in the mind, not in objects . . . (*ibid.*, p. 165).

This conclusion ties in with his thesis that all ideas (with the exception of the idea of an intervening shade of colour) are derived from impressions, and, except for compounded complex ideas, are merely faint copies of them. As is well known, it is this empiricist epistemology that underlies Hume's analysis of causality and induction. As impressions are discrete, disconnected sensory elements so are their faint copies or ideas, thus reinforcing the view that such cognitive units are so independent that no necessary connection can be intuited or demonstrated among them, and hence no denial of their conjunction or recurrence can be contradictory. If we could infer a necessary connection such an inference 'wou'd imply the absolute contradition and impossiblity of conceiving any thing different. But as all distinct ideas are separable, 'tis evident there can be no impossibility of that kind' (*ibid.*, p. 87). Thus Hume's impressionistic conception of nature can more accurately be described as pointilistic.

Consistent with his criterion that the contrast between ideas and impressions depends on nothing more than their vividness, Hume attributes any difference between opinion and belief or degrees of belief to their relative force or liveliness: 'an opinion or belief is ***nothing but a strong and lively idea deriv'd from a present impression related to it*** . . .' (*ibid.*, p. 105).

Granted these epistemic assumptions, it naturally follows that we cannot have any knowledge of nature as such nor any *reasonable* justification for believing in the uniformity of nature. In contrast to mathematical knowledge or deductive implications among ideas, which do exhibit necessary connections but do not constitute knowledge of nature, whatever apparent systematic connections occur among our empirical ideas or matter of fact knowledge derive entirely from subjective impressions and consequently are essentially irrational.

> Thus all probable reasoning is nothing but a species of sensation. 'Tis not solely in poetry and music, we must follow our taste and sentiment, but likewise in philosophy [including natural philosophy or science]. When I am convinc'd of any principle, 'tis only an idea, which strikes more strongly upon me. When I give the preference to one set of arguments above another, I do nothing but decide from my feeling concerning the superiority of their influence. Objects have no discoverable connection together; nor is it from any other principle but custom operating upon the imagination, that we can draw any inference from the appearance of one to the existence of another (*ibid.*, p. 103, brackets added).

Thus for Hume no objective truths or rational arguments can be adduced to support our belief in empirical inferences or scientific knowledge!

5 CRITIQUE OF HUME'S POSITION

But *is* it the case that the evidence and reasoning supporting our scientific theories are based on nothing more cogent than subjective feeling and that the extensive body of exact knowledge acquired in physics, chemistry, biology, and astronomy since Hume's day has no relevance whatsoever for our expectations and understanding of nature? Hume of course acknowledged that our ordinary behaviour and beliefs belie his sceptical conclusions, but he challenged philosophers to provide a *rational justification* for our confidence in induction, a challenge which I shall now take up.

Hume is largely correct, it seems to me, in maintaining that our everyday experience of the world does not initially or directly disclose the underlying causal powers (*i.e.*, the unobservable infrastructures, particles, forces, *etc.*) on which the manifest qualities and occurrences of the world depend. When we observe any ordinary object or scene before us, for example, we do not see the radiant energy that illuminates the visual aspects of the world, nor do we experience as such the effects of this reflected radiation on our retinas and its encoded transmission as chemical-electrical discharges through the optic nerves to the visual centres of the brain. What we experience is the outcome of this tremendously complex but necessary process, not the underlying transaction itself. Similarly, when we feel an object to be cold, hard, and metallic, we do not perceive the atomic-molecular structure of the object that accounts for these tactual properties. We know that ordinary water has a precise specific gravity and a definite freezing and boiling point, but initially we do not know anything of the molecular structure, chemical bonds, and electromagnetic weak forces or exchange of virtual photons which we now believe, on sound experimental evidence, account for these properties.

But while this is true of our initial observations—although it does not really take into account the extent to which even our primitive observations can be exploited to further our understanding of nature (*cf.* Churchland [1979])—our knowledge is not restricted, as Hume repeatedly maintains, to sensory observations and associations. In a way that no eighteenth-century philosopher—nor scientist, for that matter—could anticipate, our investigations of nature supplemented by refined instruments of observation, controlled experiments utilizing ever more sophisticated apparatus, physical and chemical analysis, mathematical extrapolation and theorizing have enlarged our knowledge tremendously. Hume not only repeatedly affirms that certain explanations which are commonplace today were not accessible then, he asserts that they *never* could be attained, the phenomena being so 'incomprehensible'. Citing his own examples, the impact of billiard balls, the production of a sound by a vibrating string, the occurrence of 'palsy', the control we exercise by means of our nerves over the muscles of the body, and the digestion of bread, not only can such phenomena not be explained, according to Hume, but any explanation is intrinsically 'inconceivable'—as in fact it was in terms of his epistemology! As he says in the *Enquiries* ([1777], p. 72) regarding the cause of movement by physical impact: 'We are ignorant . . . of the manner in which bodies operate on each other: Their force or energy is entirely incomprehensible. . . .' Concerning the production of sound by a vibrating string, he claims that apart from the fact that '*this vibration is followed by this sound . . . we have no idea of it*' (*ibid.*, p. 77). As for palsy and the voluntary control of our muscles by our nerves, he states that they are 'to the last degree, mysterious and unintelligible . . . wholly beyond our comprehension' (*ibid.*, pp. 66–7). But his profession of ignorance regarding the nutritive powers of bread is the best illustration of the tremendous gap in knowledge between Hume's day and our own.

> It is confessed that the colour, consistence, and other sensible qualities of bread appear not, of themselves, to have any connexion with the secret powers of nourishment and support. For otherwise we could infer these secret powers from the first appearance of these sensible qualities . . . contrary to plain matter of fact. Here, then, *is our natural state of ignorance with regard to the powers and influence of all objects* (*ibid.*, p. 37, italics added).

In another passage he says that 'neither sense nor reason can ever inform us of those qualities which fit it for the nourishment and support of a human body' (*ibid.*, p. 33).

While it is true that the nourishing properties of bread are not revealed in its sensible qualities, we are no longer limited merely to sensible qualities in explaining how and why bread nourishes. Today there is found on most packages of bread a list of ingredients such as flour, milk, and vegetable shortening, as well as nutritional components like vitamins, proteins, minerals, *etc.*, which at least partially account for the nutritional value of bread. It is the responsibility of chemists and pharmacologists in the Food and Drug Administration to ascertain by chemical analysis and metabolic studies the ingredients and harmful or beneficial properties of various foods and drugs, thereby *forewarning* the public regarding their possible effects. As disappointing as these results may sometimes be, they do constitute a form of knowledge with *predictable* consequences which Hume claimed 'never' could be attained.

6 HISTORICAL BASIS OF HIS SCEPTICISM

The essential reason for Hume's scepticism is that he believed that scientific knowledge was *inevitably* limited to the observable qualities and conjunctions of objects, or, more accurately, to our ideas of them. This belief, in turn, was founded on two factors: (1) his phenomenalistic epistemology, discussed earlier, based on impressions and ideas, and (2) the universally acknowledged limitation in Hume's day of explanations in terms of 'insensible powers'. As he frequently states, 'there is no known connection between the sensible qualities and the secret powers' (*ibid.*, p. 33). Again, however, he is not denying that 'secret powers' exist. In the *Treatise* he affirms 'that almost in every part of nature there is contain'd a vast variety of springs and principles, which are hid, by reason of their minuteness or remoteness' ([1739], p. 132), but not realising the possibility of discovering by experimentation some of these hidden 'springs and principles', thereby accounting to some extent for the properties and effects of things, he concluded that they were inevitably 'secret', 'hidden', and 'incomprehensible'. Consequently, in spite of recurrent resemblances we could have no reason for believing that the causes of these resemblances would persist and thus support our inductive inferences.

This scepticism regarding possible explanations of phenomena in terms of their underlying causes was not unique to Hume, but generally prevalent at the time. Anticipating Hume (as well as Kant), Locke, in spite of his belief in primary qualities, states that

> it seems probable to me, that the simple ideas we receive from sensation and reflection are the boundaries of our thoughts; beyond which, the mind, whatever efforts it would make, is not able to advance one jot; nor can it make any discoveries when it would pry into the nature and hidden causes of those ideas [1690].

Locke also antedated Hume in affirming that there was 'no necessary connection of real existence with any idea a man hath' (*ibid.*, IV, XI, 1).

While Locke denied any 'necessary connection between ideas and real existence', Berkeley carried the sceptical argument further in maintaining that there is no justification for inferring a 'real existence' at all, at least in terms of material or natural causes.

> I say . . . that it is possible we might be affected with all the ideas we have now, though no bodies existed without, resembling them. Hence it is evident the supposition of

> external bodies is not necessary for the producing our ideas: since it is granted they are produced sometimes, and might possibly be produced always in the same order we see them in at present, without their concurrence [1734].

If phenomena *could occur* just as they do now *without* any accompanying physical causes, then for all we know they also *could not occur as they do*, as Hume concluded. In addition, Berkeley anticipates Hume in asserting that 'food nourishes, sleep refreshes, and fire warms us . . . all this we know, not by discovering any necessary connection between our ideas, but only by the observation of the settled laws of Nature' (*ibid.*, I, 31). Hume's originality consists in denying any justification for expecting even the continuation of the 'settled laws of Nature'.

But it was probably the 'incomparable Mr Newton', as Locke described him, who was the principle source and authority for a reserved scepticism, since in a famous passage he appears to limit knowledge to a 'deduction from phenomena'—though in many other passages he attests to a deeper understanding of nature in terms of 'natural powers' and the 'forces of nature'.

> Hitherto we have explained the phenomena of the heavens and of our sea by the power of gravity, but have not yet assigned the cause of this power . . . [because] I have not been able to discover the cause . . . and I frame no hypotheses; for whatever is not deduced from the phenomena is to be called an hypothesis; and hypotheses, whether metaphysical or physical, whether of occult qualities or mechanical, have no place in experimental philosophy [1686].

Thus Hume, for all his notorious critique of the uniformity of nature and of induction, actually was following out the consequences of a well established empiricist scepticism. As such, his most eloquent expression of that scepticism, as stated in the *Enquiries*, reflected the generally accepted limits of possible knowledge of the eighteenth century.

> It is confessed, that the utmost effort of human reason is to reduce the principles, productive of natural phenomena, to a greater simplicity, and to resolve the many particular effects into a few general causes, by means of reasoning from analogy, experience and observation. But as to the causes of these general causes, we should in vain attempt their discovery nor shall we ever be able to satisfy ourselves, by any particular explication of them. These ultimate springs and principles are totally shut up from human curiosity and enquiry. Elasticity, gravity, cohesion of parts, communication of motion by impulse; these are probably the ultimate causes and principles which we shall ever discover in nature [practically a direct paraphrase of Newton's *Principia*]. . . . The most perfect philosophy of the natural kind only staves off our ignorance a little longer. . . . Thus the observation of human blindness and weakness is the result of all philosophy, and meets us at every turn, in spite of our endeavours to elude or avoid it ([1777], pp. 30–1, brackets added).

If this expression of the 'ultimate' limits of knowledge were all Hume's argument came to then I doubt that many would disagree with him; however, it is *crucial* to remember that his scepticism cuts much deeper, affirming not only that we do not have a knowledge of 'ultimate springs and principles', but that *no knowledge we can ever have* of the natures or causal powers of objects would constitute a rational justification for believing that they would continue to produce the same properties and effects (under similar conditions) in the future. Yet given the discovered fact that it is free hydrogen ions in acids that cause litmus paper to turn red, is it just as likely that in the future a liquid manifesting all the properties of sulphuric acid will turn litmus paper blue? Or considering our knowledge of the atomic-molecular structure of hydrochloric acid (HCl) and sodium hydroxide (NaOH) which when combined form salt (NaCl) and water (H_2O), is it just as possible that in the future they will combine and form potassium sulphate and mercury oxide? Or consider Popper's example [1972] of the people in a French village who died from eating their usual *baguettes*, would we entertain for one moment the belief that bread *with the very same ingredients* would have its causal powers just change in such a way as to produce those tragic effects, or would it be more reasonable to expect that some foreign ingredient had entered the bread accounting for the disastrous results, isolating ergotism as the cause?

Knowing what we do now about the function of *DNA* in cell replication, is it just as probable that when a human cell is fertilized it will develop into a giraffe, rather than a human being? Considering our knowledge of nuclear fission, when uranium atoms are split by neutrons releasing tremendous amounts of nuclear energy would it be just as likely that the world would instantly freeze over? Having constructed a match from known chemicals that will ignite under certain conditions, is it just as reasonable to expect that when we strike a match the Tour Eiffel will collapse or that the moon will change its orbit or that the holder of the match will disappear?

As far fetched as these examples are, according to Hume it is just as probable that the most familiar object will produce the most unexpected, bizzare effect because *no connection can ever be established* between the internal natures or causal powers of objects and their effects or properties. Everything is disconnected! As he unequivocally states: 'Upon the whole, there appears not, throughout all nature, any one instance of connection which is conceivable by us. All events seem entirely loose and separate' (*ibid.*, p. 74).

7 THE PRESENT UNREASONABLENESS OF THIS SCEPTICISM

But surely this conception of the world is no longer true! If in our investigations and explanations of nature we *were* limited to the manifest qualities of phenomena, such that we could never experimentally discover nor imaginatively conceive of any other domains of nature, then Hume's conclusion would be plausible. If we knew nothing more than the sensory appearances of acids then of course we could not understand why they turn litmus paper red, and therefore why litmus paper is a test of acidity. Were our knowledge of hydrochloric acid and sodium hydroxide restricted to sensory observations then we could not know why they form salt and water when combined. If we had no knowledge of atomic or nuclear physics, then how could we produce a controlled nuclear reaction? Knowing nothing of the ingredients of bread and why they have the physiological effects they do, we would be ignorant of why it nourishes. If only the external body could be observed and therefore we knew nothing of the difference between striated and unstriated muscle fibres and even less of neurons and the transmission of electrical discharges to the muscle tissues, then of course it would be 'incomprehensible' as to how our nerves activate our muscles. What would be the purpose of building electron microscopes, cyclotrons, linear accelerators, and giant optical and radio telescopes with the intent of penetrating more deeply into the inner core and outer recesses of the universe, if knowledge were limited soley to observable phenomena? But even more conclusively, how could we have landed astronauts on the moon, with all the sophisticated technology, exact understanding of forces such as gravity and thrust, and precise calculations and predictions of manoeuvres that this required, if our scientific knowledge were as limited as Hume maintained?

If the Humean persists in asking why these discovered causes, structures, and functions have the effects they do, then one can only reply that an explanation has to be in terms of something (unless we eventually arrive at a bootstrap theory). One can ask for deeper explanations, but even these deeper explanations eventually would have to terminate in the fact that some things or states (such as fields) just have the properties and effects that they do. Yet this *does not* concede Hume's thesis. Hume claimed that since we could only observe disconnected sensory qualities or ideas we could never know the 'hidden natures' or 'secret powers' that produce the observable similarities and sequences which he never denied. If we now respond to his scepticism by replying that these observable effects are produced by experimentally discovered and confirmed substructures and the Humean persists in demanding, 'but why should these substructures have these effects', then we need only reiterate that an explanation will always have to be in terms of something, otherwise we would be led to an infinite regress, which was not Hume's argument.

Or, we can take the offensive and ask the Humean what would be an acceptable answer? If explaining acidity in terms of hydrogen ions, chemical combination in terms of sharing free electrons, digestion in terms of nutritional value, solidity in terms of strong and weak nuclear forces or the exchange of virtual particles are not explanations, then what would be? If landing men on the moon is not an indication of confirmed inductive generalizations and predictions, what would be accepted as confirmation? A question for which there is no possible answer is not a reasonable question!

Consider the example of a spring driven, stem watch. If a person were unfamiliar with the works of such a watch he could observe the regular movement of the hands but would not understand how or why they rotate: though the watch kept perfect time he would have no reason for expecting it to continue to keep time in the future. But suppose one explains to the person the mechanism of the watch in terms of springs, gears, jewels, stem, *etc.*, and also removes the base of the watch exposing the works. Would he now have more reason, based on his understanding of the functioning of the mechanism, to expect that when wound up the watch will continue to run in the future? I think we would all agree that he would. But suppose the Humean persists and asks, 'why should we expect, given the observed mechanism, that it will continue to function in the future as it has in the past?' Supplementing the previous explanation, one can answer in terms of the molecular properties of metals and Newton's laws of dynamics, explaining the tension of springs, the engagement of gears, the hardness of metals, the crystalline structure of jewels, *etc.*, and the way in which all of these components interact under the usual conditions of gravity, humidity, and temperature to produce the movement of the hands. If the Humean still demands, 'why should we expect the atomic-molecular structures and dynamic laws to persist so that they have the same effects in the future,' how are we to reply? What would the Humean accept as an answer? The point is that a reasonable answer to Hume's scepticism has been provided and any further scepticism is otiose. Knowledge of the world as we know it is possible precisely because phenomena, although existing within certain contexts or conditions, possess a certain degree of autonomy so that we can explain the occurrence of such phenomena relative to those conditions without possessing a complete or ultimate knowledge of nature (*cf.* Schlagel [1981]).

8 CONCLUSION

Quite simply, then, since Hume's scepticism regarding the uniformity of nature and of induction depended upon a lack of knowledge as to how and why things work and produce the effects they do, the resolution of this scepticism consists in demonstrating that today we do possess such knowledge, at least to a certain degree. By examining phenomena under various experimental conditions (*e.g.*, Rutherford's bombardment of gold foil with alpha particles to determine its atomic structure to Watson and Crick's investigations of phages and X-ray diffraction patterns of molecular structures to determine the precise molecular structure of *DNA*) in order to discover their 'hidden causes' and 'secret natures', and devising theories to interpret the experimental results, we have gained considerable insight into the workings of nature. The greatest success to date has been achieved with the atomic-molecular theory, as evidenced by the Periodic Table, chemical analysis, atomic, nuclear, and particle physics, molecular biology, and spectroscopic analysis in astronomy. While it is undoubtedly true that we are infinitely far from any final, conclusive, or 'ultimate' explanation of most natural phenomena—if such an account is even possible—this does not mean that the knowledge we do possess has no explanatory significance or predictive value. Fortunately, we do not have to know everything to explain something. When we can determine the inner composition of substances and isolate the causes of phenomena so as to explain, control, and predict natural occurrences, even to the extent of preventing or curing diseases and creating new subnuclear particles, it would be *unreasonable* to disbelieve in our ability to discover some of the underlying conditions and physical interdependencies that support our inductive inferences. *Where experimental discoveries and theoretical constructs and frameworks have supplemented our limited and fragmentary sensory observations, our knowledge is more complete and our belief in the deeper interconnected causal matrix producing nature's manifestations more justified.*

In fact, such discovered interconnections attest both to a *natural* and a *conceptual* form of necessity, although because they are empirically discovered they are *a posteriori*, rather than *a priori* (*i.e.*, discoverable by reason itself), in contrast to the convictions of Hume and Kant. As Harré and Madden [1975] state in their excellent critique of Hume's position and defence of 'Causal Powers':

> The natural necessity in the world is reflected in a conceptual necessity in discourse about the world. Predicates are bound into ensembles by virtue of the joint origin of the properties they ascribe to things, in the nature of those things. When we think there is a natural necessity between manifested properties and hypothesised dispositions, that is, a real connection via the nature of the thing, then we are entitled to make a conceptual link, incorporating the power or tendency to manifest the property within the concept of the thing or substance.

When substances and elements can be defined in terms of more basic particles, fields, or forces which also account for their physical properties, physiological functions explained by microstructures and chemical-electrical processes, and natural occurrences traced to necessary operative conditions, nature reveals a type of necessary connections that becomes embedded in our concepts and theories. These concepts and theories in turn provide conceptual implications from which necessary inferences can be made. Predictions in nuclear physics of particles whose existence was only confirmed years later, such as the neutron, neutrino, positron, and possibly quarks, confirm this. In addition, our theories provide more reliable criteria for identifying phenomena than those derived from sensible qualities. A solution (of sufficient strength) that does not turn litmus paper red is not an acid; a metal that does not have a density of 8.9, an atomic weight of 63.5 and number of 29, along with the typical properties of malleability, fusibility, ductility, and electrical conductivity is not copper; a liquid possessing the usual macroscopic properties but that does not have the molecular structure (or is not an isotope) of H_2O would be an anomaly.

Our concepts of things and their properties, as well as our broader theoretical interpretations, mirror the experimentally discoverable interconnections in nature. Though it is we who create the concepts and theories, they are based on controlled observations and carefully designed experiments to ensure as tight a fit with nature as possible. In words echoing those of Harré and Madden, Newton-Smith [1981] says:

> Certainly it is the way the world is that makes it true that water is H_2O. But what makes this truth necessarily true is not the way the world is, apart from our linguistic practices. For the necessity is a reflection of a general institutionalized linguistic practice of determining the extension of natural kind words by reference to the nature of ostended paradigm instances of the kind in question. The necessary connection between being water and being H_2O is imposed by us on the world. For given that water is H_2O our practice is not to count as water anything that does not have this nature.

Though similar to the position of Harré and Madden, in attributing the necessity of connections more to our conceptual-linguistic framework than to nature, Newton-Smith's position is more like the 'pragmatic *a priori*' of C. I. Lewis (*cf.* [1929]).

Accordingly, our framework of knowledge and network of beliefs do not consist of disconnected, contingently associated ideational or semantic elements that have no referential or representational significance as far as the world is concerned and therefore are completely devoid of predictive value. Although our theoretical frameworks are undoubtedly 'underdetermined by empirical evidence', as Quine [1960] claims, and therefore *are subject to constant revision,* they do constitute a systematic interpretation of certain domains, at least, of physical reality enabling us to draw certain inferences and make verifiable predictions. It would not have been possible to predict and confirm the existence of such previously unknown entities as the planet Neptune and the legionnaire's virus, or establish physical equivalences such as light and electromagnetic radiation and mass and energy, if our theories did not impinge upon physical reality to some degree.

Even the statistical predictions in quantum mechanics, in so far as they are precise predictions within certain parameters, represent objective constraints on quantal events. And though the quantum formalism does not provide a classical picture of the subatomic world, few scientists would maintain that there is no connection between successive observations or interactions with the apparatus and the superposition of states in the developing quantal system as prediced by Schrödinger's wave equation or quantum electrodynamics. As limited and tentative as our knowledge of nature is, in the sense that we may never attain an *exhaustive* knowledge of the *ultimate* natures of things (and here I may disagree somewhat with Harré and Madden), it is not illusory. So if we can be relatively certain of anything, it is that the human condition is not the Humean condition.

The George Washington University

REFERENCES

BERKELEY, G. [1734]: *The Principles of Human Knowledge,* second edition, Part I, sect. 18.

CHURCHLAND, P. M. [1979]: *Scientific realism and the plasticity of mind,* ch. 2. Cambridge University Press.

DESCARTES, R. [1642]: 'Meditation III,' second edition, in *The Philosophical Works of Descartes*, volume I, E. S. Haldane and G. R. Ross (*eds.*), p. 162. Cambridge: At the University Press, 1967.

HARRÉ, R. and MADDEN, E. H. [1975]: *Causal Powers*, p. 6. Basil Blackwell.

HUME, D. [1739]: *A Treatise of Human Nature*, L. A. Selby-Bigge (*ed.*), *Hume's Treatise*. Oxford: At the Clarendon Press, 1888.

HUME, D. [1777]: *Enquiries*, posthumous edition, L. A. Selby-Bigge (*ed.*), *Hume's Enquiries*, second edition. Oxford: At the Clarendon Press, 1902.

KANT, I. [1783]: *Prolegomena to Any Future Metaphysics*, L. W. Beck (*ed.*), p. 5, brackets added. The Bobbs-Merrill Company, 1950.

LEWIS, C. I. [1929]: *Mind and the World Order*, ch. VIII. Dover Publications, Inc.

LOCKE, J. [1690]: *An Essay Concerning Human Understanding*, Bk. II, ch. 23, sect. 29.

NEWTON, I. [1686]: *Principia*, Volume II, pp. 546–7. University of California Press (English translation by A. Motte, 1729).

NEWTON-SMITH, W. H. [1981]: *The Rationality of Science*, p. 170. Routledge and Kegan Paul.

POPPER, K. [1962]: *Conjectures And Refutations*, pp. 54–5. Harper and Row Publishers.

POPPER, K. [1972]: *Objective Knowledge*, p. 11. Oxford: At the Clarendon Press.

QUINE, W. V. [1969]: *Ontological Relativity and Other Essays*, p. 72. Columbia University Press.

QUINE, W. V. [1960]: *Word and Object*, p. 78. The Technology Press of M.I.T. and John Wiley and Sons.

SCHLAGEL, R. [1981]: 'Contextualistic Realism,' *Philosophy and Phenomenological Research*, volume XLI, no. 4, pp. 437–51.

STRAWSON, P. F. [1958]: *Philosophical Studies*, volume 9, no. 1–2, p. 20*f*.; cited from Popper [1972], p. 11.

Brit. J. Phil. Sci. 35 (1984), 375–380 *Printed in Great Britain*

Discussions

O'HEAR ON AN ARGUMENT OF POPPER'S

In his study of the philosophy of Karl Popper, Anthony O'Hear in his [1980] notes (p. 145) an argument of Popper's in his [1972] (pp. 223–4) to the effect that 'there is something self-defeating about asserting that physical determinism is true, because this would be tantamount to an admission that our acceptance of determinism is not really based on our perception of the logical cogency of the arguments, but on the physical condition which makes us deceive ourselves into thinking that we are swayed by the truth'. This argument, O'Hear avers, 'the determinist could surely counter by saying that there is no reason why some people (or some machines) should not be determined in such a way that they accept only those positions which are reasonably argued . . .'.

Now I want to show that O'Hear's suggestion about how the determinist could counter Popper's argument will not do. Let us note, first, that the determinist's counter-argument does not imply that some people would 'accept only those positions which are reasonably argued' because of their perception of the logical cogency of the arguments supporting those positions; rather, they would accept them because they were *determined* to accept them. Determined by what? In accordance with the principles of physical determinism, they could be determined only by physical causes. Thus, on the determinist's counter-argument, some people would accept conclusions only of reasonable arguments, without, however, apprehending the reasonableness of the arguments establishing those conclusions. Specifically, some people might believe physical determinism to be true, but not because of their 'perception of the logical cogency of the arguments' in favour of it, but because of some 'physical condition which makes' (determines) them to believe it to be true. To this extent the determinist's counter-argument does not differ from the position which Popper implies he must hold. Where O'Hear's determinist does differ from Popper is in his contention that some people (and some machines, too, but we need not consider those) might nevertheless accept only *true* conclusions of *reasonable* arguments, and hence, presumably, if they accepted physical determinism, they would be accepting a true theory which could be reasonably argued.

Now the question is, does this counter-argument of O'Hear's determinist enable him to escape Popper's charge that asserting physical determinism to be true is self-defeating? Surely not. For let us suppose that there are three people, *A*, *B* and *C*, who are 'determined in such a way that they accept only those positions which are reasonably argued' and that the rest of us, *D* . . . *N*, are not such people. Now, could *D* pick out *A*, say, as belonging to the former group? Could *D* be sure that *A* accepted only those positions which are reasonably argued? Evidently not, for in order to be sure of this he would have to be able to check to see whether the positions *A* accepts are reasonably argued. But by hypothesis he cannot do this. And he could not do this for *B* and *C* either. Hence *A*, *B*, and *C*'s acceptance of certain positions, even of only those that are reasonably argued, would be of no use as a guide to truth or reasonableness for *D* . . . *N*. But more importantly, could *A*, *B*, and *C* themselves be sure that they were among the favoured few? No, for even though they might accept only those positions which are reasonably argued, they would accept them not because of their perception of the reasonableness of the arguments on which they were based, but because of various physical causes. In other words, they could no more check to see whether their positions were reasonably argued than could any of *D* . . . *N*. And as for *D* . . . *N* themselves, they might be determined to accept positions for which no reasonable arguments could be provided—including the position that the positions which *A*, *B*, and *C* accept are *not* reasonably argued!—while yet being determined to believe the contrary. Thus on O'Hear's suggestion no one could ever tell whether any position anybody accepted was reasonably argued; and, in particular, no one could tell whether physical determinism was reasonably argued and, further, whether it was true.

But there is a further problem. The suggestion we are considering is that some people might be determined in such a way that they accepted only those positions 'which are reasonably argued'. Does this imply that there is some being who reasonably argues for certain positions, and that the people referred to accept only those positions? But who could that being be? If it itself is a being that is determined to accept only positions 'which are reasonably argued', then we get involved in an infinite regress. If it is actually capable of apprehending the reasonableness of arguments, and accepts certain positions for that reason, then evidently it is not determined in the same way as is envisaged by the physical determinist: it is not one of those people who are determined by *physical causes* to accept only those positions which are reasonably argued. At any rate, O'Hear is suggesting that even given the truth of physical determinism, it is still possible to decide that there are certain positions that are 'reasonably argued' but he leaves unanswered the question who is in a position to decide this. (One suspects that O'Hear is imaginatively viewing a closed determinist physical system from the outside and himself as the being who is capable of deciding which positions held in that system are reasonably argued. But of course if such a relationship *vis-à-vis* a closed determinist physical system actually obtained, then universal physical determinism would not be true.)

I conclude that O'Hear has not provided the physical determinist with a good counter-argument to Popper's criticism of his position.

PETER GLASSEN
University of Manitoba

REFERENCES

O'HEAR, ANTHONY [1980]: *Karl Popper: The Arguments of the Philosophers*. Routledge and Kegan Paul.
POPPER, K. R. [1972]: *Objective Knowledge*. Oxford University Press.

REPLY TO GLASSEN

The argument of Popper's in his [1972], pages 223–4, which I was criticising purports to show that we cannot rationally accept physical determinism because (i) physical determinism entails that everything that happens does so because it is necessitated by physical or natural law, (ii) hence our acceptance of determinism would be so necessitated, (iii) a rational belief is one that is accepted because it meets certain rational standards for belief, and (iv) a belief (in this case the belief in physical determinism itself) cannot be accepted both because it is necessitated and because it meets rational standards. The claim, repeated by Glassen in his [1984] is that *if* someone is physically determined to believe something (or to be in a certain belief state) then he cannot at the same time be in that belief state because he perceives the cogency of the belief. If a belief I am determined to accept is cogent, rational *etc.*, this is at most a lucky coincidence and can have nothing to do with bringing about the belief in me. That this is Popper's point is reinforced by his longer discussion of the same argument in his [1982], pages 81–8, in which he admits the possibility of deterministic calculating machines which work in accordance with the laws of logic, but denies that an individual determined to believe something could in the full sense be believing rationally. He would not be believing because he 'freely judge(d) the *arguments or reasons* in its favour to be sound'.

Popper and Glassen assert that beliefs cannot both be physically determined and accepted because they are rational. I do not base what I say in criticism of this assertion on any general doctrine to the effect that determinism and freedom of action are compatible. Indeed, as will emerge, I do not think that belief is much like free action, and I suspect that talk of belief in terms of freely judging that such-and-such, or a response to free enquiry and the like, has tended to obscure an important point about rational belief.

Let us accept for the sake of argument that neurophysiological determinism is true, and also that all mental states are also brain states. My belief states are then neurophysiologically determined brain states. But why should any of this in itself make it impossible for me on occasion to survey some of my beliefs with a view to changing those that do not appear to me to be cogent, based on evidence and so on? For my beliefs, although all physically determined states, will still be of differing quality from the point of view of their rationality, according to whether they are backed up by my possession of good evidence or not, and the rest. Now, if in this survey I realise that a belief of mine is inconsistent with evidence I have, I will thereby be disposed to give it up. The brain state that constitutes this realisation will be the cause of my giving the belief up. So my giving it up will be because of my realisation that the belief is less than cogent, *and* because of the brain state. The two things are, on the view we are assuming, the same, so it is not always a question of *either* my being in possession of rational reasons for my beliefs *or* of my having a physically determined cause for them, as Popper and Glassen think.

What perhaps lies beind Popper and Glassen's attempt to force an exclusive disjunction where determinists might wish to see an identity is the feeling that a determinist would be unable to make a valid distinction between a belief held because it meets a reasonable standard and one held on non-rational grounds, upbringing, indoctrination, prejudice and the like. Popper says that physical determinism leaves no room for the difference between brainwashing and learning, on the grounds (I think) that the presumption of truth and rationality implicit in our use of argument cannot be explained deterministically or materialistically. Glassen implies that we could not explain a notion such as reasonableness from within a closed physical system. Now, I simply fail to see why someone should not (again be determined to) survey and examine the types of grounds on which people in his community hold their beliefs. Some will appear to lead to truth (correspondence with what is actually the case) more often than others. This person, let us say he is Glassen's *D*, would now be in a position to decide that some people (*A*, *B*, *C*) invariably held beliefs only on these truth preserving grounds, while others, *E* . . . *N*, and *D* himself, sometimes failed in this respect.

Glassen would, I think, reply at this point that the grounds on which *D* decided that *A*, *B* & *C* used truth-preserving grounds, for their beliefs, while *E* . . . *N* and he himself sometimes failed to do this would themselves be one he was determined to accept, and hence not rationally grounded. *D* would have to know what *was* true in order to distinguish between the 'good' strategies used by *A*, *B* and *C* and the 'bad' ones used on occasion by *D* . . . *N*, but all he can really know is what he is determined to think is true. *D* cannot know that the positions and beliefs *A*, say, accepts are reasonable or correct.

Now, while I accept that all *D*'s mental states are *ex-hypothesi* determined, there is no reason on deterministic grounds why *D* should be enclosed within an unbreakable circle of what he individually and idiosyncratically is determined to accept as true. He could (be determined to) select a set of statements, such as observational statements, on which there was community-wide agreement, and taking them as a basis of truth, examine which members of the community were most reliable in what they predicted would happen, before it did, retrodicted what had happened without themselves having direct knowledge of it, and so on. Recognising a community-wide set of basic truths, and recognising that the rest of the community corroborated him on the choice of such a set, *D* could then set about distinguishing those of his fellows (including himself) who do and those who do not use strategies for belief, perceptual and inferential, that generally and reliably checked out against the universally accepted set of truths. These strategies will determine what is to count as reasonable or unreasonable in belief.

An underlying point on which I disagree with Popper and Glassen is on whether there are insuperable problems with thinking of people being physically determined to be in states whereby they examine and alter their beliefs and standards of evidence in the light of their truth and rationality, and whether the possession of these truth inducing standards themselves might not be given some sort of physicalistic explanation, say in terms of their general usefulness to the species in general and to those who pay particular attention to them in particular. But why might these possibilities not obtain? People who were physically determined to examine and to respect standards of reasonableness in belief would tend to have some advantages over others; and they would also arrive at some or most of their beliefs because of the neurophysiological state they were in that constituted this respect for reasonableness. It would then not simply be a matter of chance that their beliefs were in accordance with the evidence. Of course, we are looking at things here from a deterministic point of view, there remains a sense in which it would not be up to us whether we were rational, respectful of truth and so on, and there may be objections to this arising from other considerations. But what Popper and Glassen are attempting to show is not that determinism makes it impossible to say that our beliefs are rational in some strong sense in which we are ultimate masters of our fate, but that it makes it impossible to mean anything by speaking of belief held rationally or irrationally. I have been urging that this distinction—belief held rationally or not—could survive physical determinism, because one could still be physically determined to distinguish between beliefs held on good reasons and those held on bad reasons, and indeed be physically determined to hold some precisely because they were backed by good reasons, and to realise this. There thus seems nothing inconsistent in someone claiming that his belief in physical determinism is itself both determined in him and that it meets whatever standards of rational belief obtain in the area.

In fact, as I said earlier, belief does not operate like free action: one significant difference is that the more rational a belief is and the more overwhelming the evidence in its favour, the less one would have any semblance of choice or discretion in regard to it. Descartes, whom Popper adduces as an ally in his argument against determinism, actually regards total rationality in respect of a belief as consisting in that full conviction that arises from the impossibility of believing otherwise, when confronted with a genuinely clear and distinct idea. Being compelled or necessitated to believe something, then, it not necessarily a symptom of irrationality in one's belief. Indeed, if the necessitation is necessitation of the right sort, quite the opposite is the case.

The physicalist can surely maintain that sometimes necessitation of the right sort is equivalent to one's belief having the right causal ancestry, and ultimately the right neural link-ups with the world. Popper and Glassen doubt whether the physicalist is entitled to speak of some causal ancestries being more appropriate than others, but so long as rightness is spelled out in terms of types of ancestries more likely than others to produce true beliefs, defined in terms of generally acceptable basic truths, there seems no reason to accept this. Of course saying that the determinist has the grounds available to make a distinction does not mean that he is necessarily able to apply it correctly in every case. It may be that he is determined to believe, wrongly as it happens, that his belief in, say, determinism is rational. But the case of Popper and Glassen depends on the determinist being unable to mean anything by calling a belief rational, and that they have not shown so long as the determinist is able to distinguish in the causation of belief between causation that is likely to be relevant to the truth of the belief and the causation that is irrelevant to it. The determinist must represent himself as determined to believe whatever he believes, but he may still be able to distinguish even in his own case between beliefs which had been caused in him by factors relevant to their truth and those which had been caused by irrelevant factors such as brainwashing, indoctrination and prejudice, and he may intelligibly (if not necessarily correctly) regard his belief in determinism as being of the former sort.

ANTHONY O'HEAR
University of Surrey

REFERENCES

GLASSEN, P. [1984]: 'O'Hear on an Argument of Popper's', *The British Journal for the Philosophy of Science*, 35, pp. 375–77.
POPPER, K. R. [1972]: *Objective Knowledge*, revised edition, 1979. Clarendon Press.
POPPER, K. R. [1982]: *The Open Universe*. Hutchinson.

Brit. J. Phil. Sci. 35 (1984), 381–401 *Printed in Great Britain*

Review Articles

THE EPISTEMOLOGY OF EXPERIMENT*

At the risk of sounding like a secondary school student writing his first book report I suggest that Ian Hacking's ***Representing and Intervening*** should be required reading for all of us who are seriously concerned with the philosophy of science. The book is divided into two sections. The first, 'representing', is concerned with the realism of physical theories and is both an excellent summary of and commentary on contemporary discussions of this issue, and will be discussed briefly below. The second, 'intervening', is the first book length discussion of the philosophy of experiments since Francis Bacon, and it is this section which makes the book so significant.

As Hacking states, experiments have been neglected by historians and philosophers of science for far too long. So much so that the distinguished historian and philosopher of science, Thomas Kuhn, who has written a three hundred and fifty page book on the origins of quantum theory, ***Black-Body Theory and the Quantum Discontinuity*** [1978], makes almost no mention of the experiments on black-body radiation which caused the discrepancy between theory and experiment and led to the suggestion of the quantum. The title of the book is itself significant. The emphasis, as in most work in the history and philosophy of science, is on theory. We never see a graph of the data and how it compared with theory, or find a discussion of what the experiments were and how the data were acquired. Lummer, Pringsheim, Rubens, and Kurlbaum are at best peripheral characters. One might even imagine a philosophical or historical Tom Stoppard writing a play 'Lummer and Pringsheim are Dead'.

Even an historical study of the Michelson-Morley experiment (Swenson, [1970]) fails to point out that in the 1887 paper there is an important difference between the raw data given in the tables, which gives a large linear drift, and the graph of residuals, which gives the well known null result. Michelson and Morley do not, in fact, describe the method by which they went from the data to the residuals. For an excellent study of this point see Handschy [1982].

In this review I will concentrate on Hacking's discussion of experiments.[1] As he points out, and is illustrated above, the neglect of this subject makes anything written about it novel. But Hacking has given us far more than just novelty. The questions he raises about experiments, his clear discussion and answers, and his cogent examples are extremely important. If in the discussion below it appears that I am critical of some of his views or suggesting additions to them, this should not be taken as lessening my admiration for this book.

---

* Review of Ian Hacking [1983]: *Representing and Intervening*. Cambridge University Press. xv +287 pp.

[1] Hacking has had the benefit of a long collaboration with the experimenter C. W. F. Everitt, and states that many of his examples as well as his approach to them are due to Everitt.

Many of these issues are brought to a focus in his chapter on microscopes (no serious pun intended). Hacking points out that most modern experiments involve complex apparatus, and so it would appear that at the very least, the results are loaded with the theory of the apparatus. Dudley Shapere [1982] has extended the idea of direct observation to include our current theoretical beliefs explicitly. In Shapere's discussion of the solar neutrino experiments he states, 'There is no way known other than by neutrinos to see into a stellar interior.' He further analyses this as follows '*X* is directly observed if (1) information is received by an appropriate receptor and (2) that information is transmitted directly, *i.e.* without interference, to the receptor from the entity *X* (which is the source of the information.)' The dependence of 'seeing' on current theory is clear.

Hacking points out that while one needs a theory of light to construct or improve a microscope or to design a new and different type of microscope, one doesn't need such a theory to 'see' through a microscope. His discussion suggests that I, most biologists, and probably most of his audience, have an incorrect view of microscope theory. Most of us believe that the microscope is enhanced normal vision, depending on the transmission, reflection, and refraction of light. It was first shown by Ernest Abbe, in 1873, that a microscope works by the diffraction of light. Despite the incorrect theory both we and observers before Abbe's work had confidence in images produced by a microscope. In fact, perfectly good microscopes were constructed earlier, and even when Abbe joined the firm of Carl Zeiss, which became a leader in microscope construction, it was some time before their microscopes had better resolution than those produced by English and American instrument makers, who resisted the theory and worked by purely empirical experience. Hacking notes that microscope images are robust despite such theory changes. In his view, we learn to see through a microscope not by understanding the theory, or even by just looking, but by doing, that is by manipulating the things we see. One can stain, inject fluid, and in other ways act on the objects seen.

The question then arises as to whether what one observes is an artifact of the apparatus or the observing process, or is, in fact, *there*. Hacking states that the fact that the same patterns show up in different microscopes, and there are a plethora of them: ordinary, polarising, phase-contrast, fluorescence, interference, electron, and acoustic-argues for their real existence. Hacking states that it would be a preposterous coincidence if the same arrangement of dots were produced by two totally different kinds of physical systems.

One may question whether 'different' here is theory laden. While one certainly has a rough and ready idea that the microscopes which use various properties of light, *i.e.* polarisation or interference, or different objects such as electrons or sound waves, are different, it seems that it is our theory of light that makes us consider them different. In some recent work, Colin Howson and I [1984] have discussed why 'different' experiments give more confirmation of an hypothesis than repetitions of the 'same' experiment. We formally define different experiments e and e' as those for which $P(e'_{k+1}|e_1\Lambda e_2\Lambda\ldots\Lambda e_k) < P(e_{k+1}|e_1\Lambda e_2\Lambda\ldots\Lambda e_k)$ and vice versa. We argue that at least one reason for making such a probability assignment is that the theories of the apparatus, which can be differentiated by looking at their consequence classes, are different. When the theory is the same one can point to differences in the size, shape, geometrical arrangement *etc.* of the apparatus. Theory-laden or not, Hacking's argument holds, I believe.

The careful reader will also note that the argument from 'different' microscopes bears a similarity to the 'inference to the best explanation' argument for realism. Hacking argues persuasively that it is not. The argument is much more localized. The dots in the micrograph are not explained by any theory but by the existence of an entity, so that it is really an inference to a cause. It is something real which gives rise to the dots on the micrograph.

There is, however, a related question which Hacking does not discuss in any detail: namely how do we validate observations that can be made by only one technique. Although the same dots show up in both fluorescence and electron microscopy and this argues for the reality of the dense bodies, one of the major values of electron microscopy is that it shows the existence of objects unobservable by any other technique. How then do we regard these objects? One may, of course, argue that we have confidence in the electron microscope because where it can observe the same things as other techniques it does so. This does not seem satisfactory. Extrapolation is notoriously dangerous. Certainly the existence of a well-corroborated theory of the electron microscope would be an important part of the argument.

Two examples may help to clarify this issue. The first is the Galilean telescope. For centuries Cremonini and other Aristotelians of the early seventeenth century have been ridiculed for their refusal to look through Galileo's telescope to observe the moons of Jupiter. Even some of those who did denied their existence. Their scepticism is not without some merit. Galileo's telescope was not very good, and it was only his own experience using it which enabled him to observe the moons. Even granting the observation of specks of light, how do we assert their real existence as moons of Jupiter and not as an artifact created by the telescope. One cannot here resort to the theory of the telescope and its independence of the distance of the object, because no such theory existed at the time of the *Starry Messenger*. One could argue that on earth, where one could check by direct observation, the telescope seemed to give true images. But the extrapolation to astronomical distances is of many orders of magnitude. I suggest here that it is the observed phenomena themselves which argue for their validity. Although one might imagine that the telescope could create specks of light, it hardly seems possible that it would create them so they appeared to be a small planetary system with eclipses and other consistent motions. It is even more preposterous to believe, that they would satisfy Kepler's Third Law ($R^3/T^2 = \text{constant}$), although this argument would not have been available until Kepler's publication of his *Harmonices Mundi* in 1619.

A similar problem arises with respect to the radiotelescope. Although the radiotelescope can detect some, but not all, of the sources visible to an optical telescope, its major interest is in the detection of sources not visible in such telescopes. Although the overlap of sources visible to both telescopes argues for both the existence of the sources and the usefulness of the two telescopes, this seems insufficient to establish the validity of radiotelescope observations. I would suggest here that at least part of the argument is theoretical. First there is a well-corroborated theory of the apparatus. Perhaps more importantly the sources observed and their characteristics can be explained relatively straightforwardly using our current theories of astronomical phenomena. It is this theoretical explanation which provides at least some of the validity of the observations.[1] This is not at all to suggest that the phenomena are observed because they are predicted, as some might have us believe, but that the theoretical explanation can give us confidence in our observations.

Hacking emphasises, however, quite rightly that one must be careful to distinguish between the theory of the apparatus and the theory of the phenomena. He discusses three possible ways to investigate Dicke's hypothesis that the interior of the sun is rotating: (1) Use optical observations of the oblateness; (2) measure the sun's quadrupole mass-moment with a satellite fly-by; (3) measure the relativistic precession of a gyroscope in orbit around the sun. Only the third method is independent of our theory of the sun and is really an observation of the interior rotation of the sun. The other two involve assumptions about the sun and thus really only suggest inferences. In the experimental discovery of the nonconservation of parity in the weak interactions it was very important that the apparatus used was, in fact, independent of the weak interactions and did not assume such nonconservation for its proper operation (Franklin and Smokler, [1981]). A more trivial example might be the use of a mercury thermometer in the investigation of whether an object's volume increases with temperature. Here the apparatus itself depends on the phenomenon to be investigated. Clearly another kind of thermometer is to be preferred.

I suspect that finding any general model of how we come to believe justifiably in experimental results will be extremely difficult, if not impossible. As the above examples illustrate, the reasons for such belief vary with the particular experiment considered. I would suggest, however, that any such model will include Hacking's emphasis on the use and manipulation of the apparatus and the use of different kinds of apparatus to observe the same thing. In addition, as noted above, I believe the model will include, at least some of the time, both the theory of the apparatus itself and also the theory of the phenomena observed.

---

[1] I am grateful to my colleague, Professor James Warwick, a distinguished radioastronomer, for helpful discussions of this point.

At least some of the difficulties of such discussions of the validity of experimental results are illustrated by the recent positive results on the observation of fractional charges (not integral multiples of *e*, the charge on the electron) reported by Fairbank and his collaborators (LaRue, [1981]). (For an interesting discussion of this experiment up to 1979, see Pickering, [1981].) Following the suggestion in 1964 of fractionally charged quarks as the fundamental constituents of all strongly interacting elementary particles various attempts were made, using many different techniques and apparatuses, to detect such particles. The consensus of the physics community that these searches had failed to find these quarks led to the *ad hoc* introduction of quark confinement into the empirically successful theory of strongly interacting particles, quantum chromodynamics, or *QCD* for short. Although such experimental searches have continued, none have been successful, with the exception of the Fairbank experiment. One difficulty here is that Fairbank's apparatus which uses magnetic levitation of superconducting niobium spheres and *SQUID* (superconducting quantum interference device) detectors is at the frontier of experimental techniques. Its theory of operation and its possible artifacts are not well understood so that his results do not command universal, or even significant minority, assent. Another problem is that the major competing experiment, that of Morpurgo and his collaborators, using similar, although not identical apparatus with carbon and iron spheres, gives negative results, although it is not conclusive.[1] The anomaly of Fairbank's results is further emphasised by the fact that recent approximate *QCD* calculations give confinement. Although the consensus of the community is that Fairbank's results are artifacts, no one has yet offered a satisfactory explanation of how such artifacts can arise.[2] The issue is still unresolved, but the problem of the validity of experimental results remains clear.

The discussion above bears on another point made by Hacking concerning the repetition of experiments. He points out that no one ever really repeats an experiment (*pace* Heraclitus) but intends to do it better. While this may be the intent of the experimenters in many cases, a quick glance at the Table of Particle Properties published regularly in *Reviews of Modern Physics* shows clearly that the same quantity, particularly those with theoretical significance, is often remeasured. Sometimes the repetition is to establish the result and at other times it is to look for the result under varying conditions. Hacking notes, as discussed above, that such a variety of experiments makes the result more compelling. This is, however, true only if the experiments give the same results. When, as in the case of Fairbank and Morpurgo, they disagree, a problem arises. In some sense they are doing the same experiment, searching for fractional charges. The differences in apparatus, particularly the use of niobium as opposed to carbon and iron, are important. What is needed is a repetition of Fairbank's experiment, using nearly identical apparatus and technique, to establish that the effect is real. The currently planned searches for fractional charge by rather different methods may convince the physics community that free fractional charges do not exist, but the anomaly of Fairbank's results will still persist.

---

[1] Morpurgo's experiment does, in fact, occasionally give fractional charges. These are eliminated by separating the capacitor plates. Unfortunately, Morpurgo does not use the same technique on spheres which give integral charges, to see if fractional charges appear.

[2] Fairbank's results are stabilised using their output. In order to guard against possible psychological bias in settling on a fractional charge Luis Alvarez suggested that a random number be added to the output so that the experimenter does not know what value is being stabilised. Fairbank is doing this but, as yet, no results have been reported. In a recent talk at the San Francisco meeting (Nov. 1983) of the American Physical Society the group at San Francisco State University used this technique, in a similar experiment on gold and mercury, and found no fractional charges. While it may be regarded as wildly implausible by the physics community, one may speculate that recent theoretical results on fractional charges in one-dimensional systems (Su and Schrieffer, [1981]) might some day explain Fairbank's results as an artifact. If such effects exist in one-dimension it is possible that they may exist in a two-dimensional system such as the surface of a superconducting niobium sphere.

Hacking also offers impressive evidence against the view that all observations and experiments are theory-laden. This view is forcefully stated in Popper's *Logic of Scientific Discovery* [1934].

> 'The theoretician puts certain definite questions to the experimenter, and the latter by his experiments tries to elicit a decisive answer to these questions and to no others. All other questions he tries hard to exclude. . . . It is a mistake to suppose that the experimenter [. . . aims] "to lighten the task of the theoretician," or . . . to furnish the theoretician with a basis for inductive generalizations. On the contrary the theoretician must long before have done his work, or at least the most important part of his work: he must have formulated his questions as sharply as possible. Thus it is he who shows the experimenter the way. But even the experimenter is not in the main engaged in making exact observations: his work is largely of a theoretical kind. Theory dominates the experimental work from its initial planning up to the finishing touches in the laboratory (p. 107).'

In 1959 Popper added a footnote stating that he should have also emphasized, 'The view that observations, and even more so observation statements, and statements of experimental results, are always ***interpretations*** of the fact observed; that they are ***interpretations in the light of theories.***'

Hacking points out that the relationship between theory and experiment is much more complex. It differs at different stages in the development of a science, nor do all the natural sciences go through the same cycles. Hacking presents us with a marvellously illustrated taxonomy of experiment and observation ranging from what he calls pristine observations such as William Herschel's observation of 'radiant heat' and Caroline Herschel's discovery of comets to the massively theory-laden experiments on the interior rotation of the sun and on solar neutrinos, discussed above. Hacking gives an interesting counterexample to Popper's view in Davy's observation of the gas emitted by algae and the flaring of a taper in this gas. Davy had no theory. If he had subsequently said, 'Ah, then it is oxygen,' he would have been making an interpretation but he did not do so.

Hacking's taxonomy includes 'noteworthy observations' such as Bartholin's observations on Iceland Spar, Grimaldi's and Hooke's observations of diffraction, and Newton's dispersion of light. Hacking states, 'Now of course Bartholin, Grimaldi, Hooke, and Newton were not mindless empiricists without an "idea" in their heads. They saw what they saw because they were curious, inquisitive, reflective people. They were attempting to form theories. But in all these cases it is clear that the observations preceded any formulation of theory (p. 156).'

Another class of experiments are those which give rise to phenomenological laws and await theoretical explanation. Hacking provides us with an example from solid state physics. In their then standard textbook, the theory of the *Properties of Metal and Alloys* [1936], Mott and Jones give several experimental results that any decent theory of metallic conduction of heat and electricity must cover. These include: (1) The Wiedemann-Franz law, that the ratio of thermal to electrical conductivity is equal to LT, where T is the absolute temperature and L is a constant, which is the same for all metals, (2) the magnitude of electrical conductivity for a pure metal and its dependence on its place in the periodic table, (3) the Mathiessen rule that the change in resistance of a pure metal due to small impurities is independent of temperature, (4) the dependence of resistance on temperature and pressure, and (5) the appearance of superconductivity. Mott and Jones note that quantum mechanics provided a qualitative understanding of the first four (superconductivity was explained in 1957). The point here, as Hacking shows, is that these laws were established long before there was any theory to explain them. Experimentation has many lives of its own. It sometimes stimulates theory. Hacking provides an example here in the work of Brewster on the polarisation of light which helped stimulate the wave theory of light. Brewster was not testing or comparing theories, he was trying to find out how light behaves.

There are also happy meetings of theory and experiment. Penzias and Wilson discovered the uniform 3°K background radiation while working on antennas and radiotelescopes. This meshed quite nicely with the contemporaneous theoretical work that such radiation was produced by the Big Bang. They were unaware of this work when they performed the experiment. Yet in subsequent history of the episode, by others, theory dominates. 'Radioastronomers believed that if they could aim a very sensitive receiver at a blank part of the sky, a region that appeared to be empty, it might be possible to determine whether or not the theorists were correct.' Other examples abound of such incorrect theory-dominated histories of experiment.

Although Hacking emphasises, quite properly in my view because of its neglect, the various relationships of experiment and theory, including the frequent independence of experiment, he does discuss how theory tests experiment. He points out that in both Popper and Kuhn experiment does test theory, but for Popper it is the avowed and explicit aim of experiment, while for Kuhn the anomalies produced by experiment are by-products. I think that he is perhaps too generous to Kuhn who at various times seems to suggest that if an experiment disagrees with established theory the experimenter will regard the experiment as wrong. Although such disagreement might cause an experimenter to carefully recheck his results, I believe, Kuhn underestimates the independence of experimenters and their confidence in their own results. We need only look at Fairbank's publication of his fractional charge results despite both disagreement with theory and the scepticism of experimenters to confirm this. Hacking also points out the sociological truth that in order to receive funding for the experiments, scientists must relate their results to the testing of theory. He attributes this to the fact that Popper's philosophy is now rampant, a view which his readers might question. Hacking's discussion of experiment emphasizes 'doing' rather than 'saying', but here I believe 'saying' has a real impact on 'doing'. If a scientist says that he is testing a theory, for whatever reasons, then that will have an important effect on how he states and interprets his results.

It is here that I wish Hacking had turned his critical acumen on various problems raised by philosophers concerning the testing of theory by experiment such as the Duhem-Quine problem, the tacking paradox, and the isolation of which part of a theory is being tested (Glymour, [1980]). I believe that scientists, at least in contemporary physics where the theory is mature and well-developed, do, in fact, test theory in most of their experiments. They do so, in general, happily unaware of the philosophical problems noted above. Their testing is much more localised. They test either the least corroborated part of the theory, or the part regarded as crucial for the calculation of the result. They do not test general theories. Thus for example, the $V-A$ theory of weak interactions predicts that the ratio $\Pi^+ \rightarrow e^+ + \nu / \Pi^+ \rightarrow \mu^+ + \nu = 10^{-4}$. When, as in the early experiments, although not currently, the result disagrees with the prediction it was the use of the vector and axial vector interactions alone that was questioned, not quantum field theory. A similar example concerns light scattering from CdS crystal (Leite, [1969]). Here two different theoretical calculations, one which included excitons and one which did not, were being compared. The authors conclude that excitons are important, 'This demonstration of exciton resonance further supports the theory of Ganguly and Birman....' An alternative view, that of modifying quantum mechanics is nowhere mentioned. As the principle investigator, James Scott, noted, 'Such a suggestion would have been laughed at.'[1]

[1] Private conversation. I am grateful to Professor Scott for his helpful discussions.

I have dealt here with only a few themes taken from Hacking's novel discussion of experiment and 'intervening'. There is, of course, far more, and it should all be read. The first half of the book concerns 'representing' or the problem of scientific realism. It is an excellent summary and discussion of the current debate. Hacking has a remarkable ability to present complex views in a concise, cogent, and understandable way. Although it is not a substitute for reading the original works, Hacking presents us with extremely useful summaries of Putnam, Cartwright, Smart, Van Fraassen, and Lakatos, among others. Hacking views arguments about the reality of theories as less useful than discussing experiments. These two topics are brought together in his final chapter, 'Experimentation and Scientific Realism'. His view that scientific entities are real is summed up as follows. '*We are completely convinced of the reality of electrons when we set out to build—and often enough succeed in building—new kinds of device that use various well-understood casual properties of electrons to interfere in other more hypothetical parts of nature* (italics in original, p. 265).' This is then beautifully illustrated in his discussion of PEGGY II, the source of polarised electrons, used in experiments to observe parity violation in electron scattering, and to test the Weinberg-Salam unified theory of electromagnetic and weak interactions. All through this chapter I kept thinking of a paraphrase of Marx, with which Hacking would agree. 'Theorists have only given us various ways of understanding electrons. The point, however, is to use them.'

In summary, Hacking's book is an extremely valuable contribution. It is subtitled, 'Introductory Topics in the Philosophy of Natural Science.' But, as Hacking states, introductory does not mean simplified. It is an excellent book and its availability in a paperbound version makes it readily available for students. I urge you all to read it and to use it in your courses on philosophy of science. I certainly intend to use it in my own.

ALLAN FRANKLIN
University of Colorado

REFERENCES

FRANKLIN, A. and SMOKLER, H. [1981]: 'Justification of a "Crucial" Experiment', *American Journal of Physics*, **49**, pp. 109–12.

FRANKLIN, A. and HOWSON, C. [1984]: 'Why Do Scientists Prefer to Vary Their Experiments?', *Studies in History and Philosophy of Science*, forthcoming.

GLYMOUR, C. [1980]: *Theory and Evidence*, Princeton University Press.

HANDSCHY, M. [1982]: 'Re-examination of the 1887 Michelson-Morley Experiment', *American Journal of Physics*, **50**, pp. 987–90.

KUHN, T. S. [1978]: *Black-Body Theory and The Quantum Discontinuity*. Clarendon Press.

LARUE, G. S., PHILLIPS, J. D., and FAIRBANK, W. M., [1981]: 'Observation of Fractional Charge of (1/3)*e* on Matter', *Physical Review Letters*, **46**, pp. 967–70.

LEITE, R. C. C., SCOTT, J. F. and DAMEN, T. C. [1969]: 'Multiple-Phonon Resonant Raman Scattering in CdS', *Physical Review Letters*, **22**, pp. 780–2.

PICKERING, A. [1982]: 'The Hunting of the Quark', *Isis*, **72**, pp. 216–36.

SHAPERE, D. [1982]: 'The Concept of Observation in Science and Philosophy', *Philosophy of Science*, **49**, pp. 485–525.
SU, W. P. and SCHRIEFFER, J. R. [1981]: 'Fractionally Charged Excitations in Change-Density-Wave Systems with Commensurability 3', *Physical Review Letters*, **46**, pp. 738–41.
SWENSON, L. [1970]: 'The Michelson-Morley-Miller Experiments Before and After 1905', *Journal for the History of Astronomy*, **1**, pp. 56–78.

NANCY CARTWRIGHT'S NEW PHILOSOPHY OF PHYSICS*

1 *The Argument for Theoretical-Law Anti-Realism*
2 *How the Fundamental Laws Explain*
3 *Theoretical Entities as Causes*
4 *Quantum Mechanics*

Nancy Cartwright has reworked ten or so of her recent essays, collected them together with some new material, and produced a brilliant book which I think will alter the course of contemporary philosophy of physics.

How the Laws of Physics Lie articulates one half of an experimentalist (perhaps one should say 'anti-theoreticist') 'Stanford' view of physics which stresses the practical activity of experimenting and the inhomogeneity of real-life physics. The other half of this new view of physics is presented in Ian Hacking's *Representing and Intervening* (Hacking [1983]) which is, as Cartwright says, the "complement" (p. 20) to her own book. The Stanford philosophy of physics recommends that we do not take *mathematical* physics 'too seriously' (p. 196). Physics is a matter of discovering the truth about physical entities, those that have 'concrete causal roles' (p. 8). Mathematics is a means to this end. But physics is not the investigation of the mathematical structure of the world. Taking mathematics too seriously can lead to trouble in the philosophy of physics: Cartwright thinks that the measurement problem in quantum mechanics is a pseudo-problem, or rather the wrong problem, and is just an artefact of the mathematics of the theory.

The essays in *How the Laws of Physics Lie* are refreshingly free of formal logic (except for Essay 1 on the topic of causal laws, the most conventional piece of work in the book). Instead, one finds circuit-diagrams for amplifiers, quite detailed discussions of laser physics, and quantum-theoretic calculations. The reader's attention is led away from the orthodox formal models of physical theories with which philosophers of science operate, and is directed instead towards the real world of physics. This is the guiding idea of the book: that the philosophy of physics ought to be the philosophy of real physics and not the philosophy of imagined physics, the philosophers' physics as it ideally ought to be. So Cartwright emphasises physics as a human activity—how it is actually done and how it is actually learned. This is an Aristotelian view of physics. It flows from 'an Aristotelian belief in the richness and variety of the concrete and particular' (p. 19). There is in *How the Laws of Physics Lie* a good deal of referring to what is written in the standard texts (texts like Merzebacher's and Messiah's *Quantum Mechanics*, and Klauder and Sudarshan's *Quantum Optics*) and a good deal of referring to what is said by science professors in class (especially in Stanford's 'Electrical Engineering 231').

---

* Review of NANCY CARTWRIGHT [1983]: *How the Laws of Physics Lie*. Oxford University Press, £16.00 (hardcover), £7.95 (paperback), 221 pp.

Parts of earlier versions of this review article were read in seminars at Balliol College, Oxford, and at the Department for the History and Philosophy of Science at Chelsea College, University of London.

This is a terse and deceptively short book. It is a measure of the book's originality that people tend to read things into it. People I've discussed it with sometimes seem to me to have read an entirely different book.

Of its many original theses the most important are I think the following.

1. *The fundamental laws of physics are not true.*

By 'fundamental laws of physics' Cartwright means '. . . Maxwell's equations, or Schrödinger's, or the equations of general relativity . . .' (p. 54).

2. *Nevertheless, the fundamental laws of physics are explanatory.*
3. *The phenomenological laws of physics are* (*or tend to be*) *true.*

'Phenomenological' is here used in the physicist's sense: the sense in which physicists speak of the phenomenology of high-energy particles or the phenomenology of superconductivity and superfluidity. Cartwright quotes 'The Encyclopaedic Dictionary of Physics': 'A phenomenological theory relates observed phenomena by postulating certain equations but does not enquire too deeply into their fundamental significance' (p. 1). So 'phenomenological' in physics and in Cartwright's book bears no simple relation to the philosophers' 'observable', around which collect the philosophers' problems which Cartwright does not think it necessary to discuss.

4. *Though not true of the physical world, theoretical laws do describe simulacra*—or models—*simplified, false, manageable constructs by means of which explanation in physics proceeds.*

('To explain a phenomenon is to find a model that fits into the basic framework of the theory and that thus allows us to derive analogues for the messy and complicated phenomenological laws which are true of it' (p. 152).) If the fundamental laws are false they cannot figure as the major premisses in deductive-nomological explanations of phenomena. And so Cartwright feels obliged to offer an alternative theory, a theory of explanation *via simulacra*.

5. *Physical theories*, or 'physics theories' as Cartwright calls them (for example on p. 77), *are not deductive structures.*

The orthodox model of physics theories as sets of sentences closed under deduction is mistaken. '. . . It is never strict deduction that takes you from the fundamental laws at the beginning to the phenomenological laws at the end . . .' (p. 104) and yet the fundamental laws *and* the phenomenological laws are parts of the theory and we do 'move' from one to other. Part of what Cartwright is saying seems to me to be this. Talk of 'physical theories' in the way that we philosophers of physics do and you are on the road to losing sight of the real object of study—real physics. Talk of 'physics theories' and you keep that object in view.

6. *Inference to the best explanation is fallacious and cannot justify belief in the truth of theoretical laws in physics*. But 'inference to the most likely cause' is not fallacious (p. 6). In its phenomenological laws physics describes causes and Cartwright's case for the existence of theoretical entities is that inference to the **entities** is justified, since these are the causal agencies which phenomenological laws describe. But inference to the 'underlying' theoretical laws is not justified.

7. Finally Cartwright develops her two recent contributions to the philosophy of quantum mechanics:

(a) the claim that *quantum probabilities are always transition probabilities* and are not, as is usually claimed, probabilities that an observable has its value lying in some range;

(b) the claim that *there is no specific measurement problem in quantum mechanics*: that the non-unitary evolution of the density operator—the collapse of the wave-packet in another idiom—is common to many quantum processes in addition to measurement.

Cartwright's position is realist on the entities of physics, but anti-realist on the theoretical laws of physics. She believes that the ontology of physics is broadly correct (that physics gets right what entities there are that have concrete causal roles) but that its attempts to arrive at the most general description of this ontology are failures. (This is the North American, strong use of the term 'realism.' One is realist about a class of sentences *iff* one holds that they are all true, rather than merely each either true or false, the usage more current on this side of the Atlantic. In this review we follow trans-Atlantic usage.)

Cartwright's final chapter on the philosophy of quantum mechanics may well appear to many readers as an afterthought, or perhaps as an application of her general philosophy of physics to a special case. But I should want to see things the other way around. In spite of the arguments she gives for theoretical-law anti-realism I should guess that it is the natural anti-realism of quantum mechanics that pulls the strings in the new philosophy of physics. Cartwright tells us the story of her conversion from a robust and tough-minded realism about Schrödinger's equation. Schrödinger's equation tells us that a macroscopic measuring apparatus (along with measured quantum systems) will generally be in a superposition of states after making its measurement. Of course measuring apparatuses appear to be determinate and therefore not generally in superpositions at all, a contradiction that seemed to Cartwright to be reconciled by the Daneri–Loinger–Prosperi theory of measurement (hereafter *DLP* [1962]). For reasons we discuss later, the *DLP* defence of Schrödinger's equation now seems untenable and Cartwright is forced into an anti-realist view of it. But once one has taken the first step towards anti-realism the rest are much easier to take, and harder to resist.

Philosophical fashion is moving strongly in the direction of anti-realism. Cartwright's book is a moderate contribution to the new wave. In the philosophy of science, one of the prime movers towards anti-realism is the stony ground that quantum mechanics is for the realist. I think we see in the development of Cartwright's philosophy of science over the last decade exactly this response to quantum mechanics: a dropping of the view that by philosophical cleverness one can give quantum theory a realist interpretation, followed by an extension to the rest of physics of the pessimism that that defeat has created, and a backing of that pessimism with independent arguments.

I THE ARGUMENT FOR THEORETICAL-LAW ANTI-REALISM

Why take the fundamental laws of physics to be false?

Cartwright's argument is this: the fundamental laws of physics are false because of the composition of forces, or more generally, because of the composition of causes as described by the theoretical laws.

First note that by the expressions 'theoretical laws' and 'fundamental laws' Cartwright always means the dynamical laws or the force laws of physics—Maxwell's equations, the Schrödinger equation, Newton's laws of motion or gravitation.

What is the argument (pp. 59–73)?

Imagine a classical charged particle moving under the influence of forces in the real world. On the one hand gravitation acts on it (because it is in the real world with masses around) and so do elecromagnetic forces (ditto an electromagnetic field). All these forces add vectorially to produce a resultant force and a resultant acceleration of the particle. But the accelerations due to each separate force do not appear to occur. So the law of gravitation and the electromagnetic laws are each false.

Stated briefly, the argument is this: because forces are vectors, force laws are false; or, *when causes combine they disappear, when forces add vectorially they too disappear and so fail to state the facts.*

This might seem at first sight an odd idea. In fact it is an idea with a proud history. Cartwright makes a passing note of J. S. Mill's discussion of much the same problem. More interestingly I think, it was the source of Russell's early philosophy of space and time and perhaps of his philosophy of mathematics as well. So, strangely, it is one of the formative ideas of the analytical tradition in philosophy. In the Preface to the *Principles of*

Mathematics of 1903 Russell writes

> A few words as to the origin of the present work may serve to show the importance of the questions discussed. About six years ago, I began an investigation into the philosophy of Dynamics. I was met with the difficulty that, when a particle is subject to several forces, no one of the component accelerations actually occurs, but only the resultant acceleration, of which they are not parts; this fact rendered illusory such caustion of particulars as is affirmed, at first sight, by the law of gravitation. It appeared also that the difficulty in regard to absolute motion is insoluble on a relational theory of space. From these two questions I was led to a re-examination of the principles of Geometry, thence to a philosophy of continuity and infinity, and thence, with a view to discovering the meaning of the word 'any' to Symbolic Logic (Russell [1937], pp. xvi–xvii).

In Russell's solution to the problem as posed about gravitation, one rejects the idea that the gravitational attraction operates between particulars and substitutes the idea that the causal relation described in the theory of gravitation holds not between particulars but rather between *configurations*—arrangements of material particulars in the universe at an instant. In fact, the causal relation is a three-place relation, holding between *three* configurations, in order that one can obtain an unique solution to the second-order differential equation of Newton's second law. The concept of configuration requires absolute space and time or so Russell argues, and this is, apparently, the source of Russell's early realism about these terms.

Cartwright's response to the problem as she poses it—about two different forces rather than two components of a given force—is in some ways similar to Russell's. She thinks that it is absurd to say that all the components of a physically real vector are also physically real (or as Russell would say, 'exist'). An argument for this view (not found in Cartwright's book): if both a pair of components *and* their resultant 'exist' the resultant would be twice the resultant and so zero; but there are non-zero resultants. Therefore, laws which describe the components of a resultant force must be false.

Cartwright's response is also similar to Mill's in some ways. (Presumably it was Mill on whom Russell had his sights.) Mill—for which see his [1868], pp. 405–6—thinks (unlike Cartwright) that both the component forces 'have their full effect'. But this principle 'by no means prevails in all departments of the field of nature'. For example, in chemistry it doesn't. 'Not a trace of the properties of hydrogen or of oxygen is observable in those of their compound, water. . . . This explains why mechanics is a deductive science or demonstrative science, and chemistry not.' Cartwright's view has something of Mill's, but only in the sense that Cartwright extends this defect of chemistry to mechanics and to physics in general. Physics, like chemistry, is not deductive in Cartwright's view.

But one can I think take another line altogether, and it happens to have been taken by Lewis Creary in his attack on Cartwright (Creary [1981]). Why not distinguish between the *mathematical* components of a force vector (of which there are non-denumerably many) and the *physically real* components? In the case of a particle subject to two forces acting in different directions the two forces are the *only* physically real forces. The resultant force is a fiction according to Creary.

Creary calls the two original force laws—of gravitation and electromagnetism—'laws of causal influence'. The resultant motion, unlike the resultant force, *is* physically real and is derived from these force laws together with a 'law of causal action' which describes how laws of causal influence combine. In the present case, the law of causal action implies the vector addition of their two separate effects.

What can be said against Creary's account? Cartwright objects, rather lamely I think, that one cannot always *find* laws of causal action for combining causal influences. In fact one can only rarely find them—'Dynamics, with its vector addition law, is quite special in this respect' (p. 63). She also objects that the laws of causal influence we do have usually contain a '*ceteris paribus*' clause which states that they are true only when no other causal influence is operating (pp. 63–7) thus ruling out the possibility of a law of causal action which would encompass them.

More serious I think is the fact that Creary's account involves rejecting Newton's second law, since there are resultant accelerations without resultant forces. Creary regards this as merely re-interpreting Newton.

> Newton's second law is interpreted as saying that the set of all natural forces acting on a body will produce an acceleration in the direction of the mathematical resultant of the forces which is directly proportional to the magnitude of the resultant and inversely proportional to the mass of the body (Creary [1981], p. 152).

Faced with these alternative views what should one say?

One should I think admit that the vector addition of forces provides a philosophical puzzle about which various postures may be taken. I confess I do not find its outcome sufficiently clear-cut to provide a basis for Cartwright's general theoretical-law anti-realism. Notice also that by 'fundamental laws' Cartwright means dynamical laws. But there are fundamental laws that seem to me to be true *tout court*. Take the exclusion principle in quantum mechanics, the requirement that the state-vector for a collection of fermions, electrons say, be anti-symmetric. The exclusion principle has the consequence that no two electrons can be in the same state. Chemistry would not be possible without it. But is it a law of physics? Or is it a 'stipulation'? If it is a mere stipulation then I find it odd that it has such an immediate and powerful consequence as the possibility of chemistry. But then I am unclear as to what a fundamental law of physics is.

2 HOW THE FUNDAMENTAL LAWS EXPLAIN

It is the aim of physics both to describe and to explain. But these are quite different functions. The fundamental laws explain but do not describe. Therefore the *D–N* or covering-law models of explanation have to go. There are two troubles with covering-law models: they let in too little, and covering laws are too scarce (p. 45).

They let in too little because, strictly speaking, there are no fundamental covering laws—true fundamental laws—and there are too few true phenomenological laws to provide adequate cover. The idea that there are plenty of covering laws involves a misunderstanding of the function of laws. Take the stock example, as Cartwright does (pp. 46–9) of Snell's law. Snell's law is the one that tells us about the angle of incidence, the angle of refraction and the refractive index. But it is true only for optically isotropic media. But '*most* media are optically anisotropic, and in an anisotropic medium there are *two* rays' (p. 47). What Snell's law enables us do is to model the anisotropic case on the isotropic one. The over-simple law of Snell enables us to explain the typically anisotropic real world of which it is false.

This is the key to Cartwright's account of explanation in physics. 'To explain a phenomenon is to find a model that fits it into the basic framework of the theory and that thus allows us to derive analogues for the messy and complicated phenomenological laws which are true of it' (p. 152). That is why we explain by ***imposing*** on Nature models from a remarkably small repertoire. Think of the way the harmonic oscillator keeps popping up in quantum mechanics (p. 145). The simplicities of the model—their 'properties of convenience' (p. 153)—are what allow us to use the mathematics of theories in explanation in physics. 'We aim to "see" the phenomenon through the mathematical framework of the theory' (p. 152). Models, simulacra, 'highly fictionalised descriptions' (p. 143) are what limit the bridge principles we use to connect mathematics with the world. But there is never goodness of fit between our mathematical theories and the complex, messy real world.

On all this Cartwright seems to me to be as near right as anyone who has written on the subject. The essential mistake of the *D–N* model is to view physics as a body of deductive arguments. But what arguments there are in physics are usually fallacious, if one adopts the logician's standards of rigour. A physicist will mix incompatible theories, appealing to this equation from one theory and that from an incompatible one. Most of applied quantum mechanics involves the use of dubious mathematical devices like perturbation theory (and methods like Padé approximants, *etc.*, *etc.*) the mathematical justification of which is always a long way behind the practical use (but perhaps Cartwright is *too* unkind to scattering theory (p. 141) which ***need*** not rely on the hocus-pocus of Dirac delta-functions). Of course, it is a mathematician's task to justify these methods and one is happier when they are justified. But the philosophy of physics is the philosophy of the physics we have (*cf.* p. 145). What else could it be? And this physics is as messy as the world (when it is successful).

We are tempted to think of the physicists' derivations as deductive arguments because of the idea that physics theories are pieces of interpreted mathematics. We reconstruct the mathematics used in physics by appealing to set theory and classical first-order logic. And so we think of physics as describing logical structures, as having models in the logicians' sense. We think we can paraphrase the language of physics into first-order logic without the problems that attend the paraphrasing of other fragments of natural language *because* physics is mathematical and already partly formalised. But this 'logicist' picture of physics cannot explain physics as an activity in which we use *contradictory theories* to flesh out a simulacrum into a description of a real phenomenon. Reconstructionism is appropriate in the philosophy of mathematics because mathematical theories do not clash. But it is wrong in the philosophy of physics where theories do (unless we are thinking of the philosophy of 'final' physics, or 'ideal' physics or 'physics as we would like it to be').

This is a corollary to another undercurrent in Cartwright's book: that theories are not sets of statements closed under deduction. I read Cartwright as conventional to this extent: that the primary objects of interest to the philosopher of physics are statements, and not (say) pictures, or other non-verbal, non-mathematical 'vehicles of thought'. But we must view theories—sets of statements perhaps—as having an open, non-deductive texture. Deriving results in physics is more a matter of knowing what approximations to make than it is a matter of knowing how to derive one statement from some others.

3 THEORETICAL ENTITIES AS CAUSES

There is a suggestion, though no more than a suggestion (p. 62), that the fundamental laws of physics may possibly describe the *causal powers* of physical systems even though they fail to describe the *facts*. But Cartwright says she is sceptical of all the current philosophical stories about causal powers and so the idea does not get off the ground. Nevertheless, if theoretical-law anti-realism is the negative thesis of Cartwright's book, *causalism* is its positive thesis. Thus 'the real content of our theories in physics is in the detailed causal knowledge they provide of concrete processes in real materials' (p. 128). But how can we be confident that physics gets its causal stories any more nearly right than its theoretical laws? Cartwright's answer is that inference to the most likely cause is not fallacious. Why? Because 'we can infer the truth of an explanation only if there are no alternatives that account in an equally satisfactory way for the phenomena. In physics nowadays . . . an acceptable causal story is supposed to satisfy this requirement' (p. 76). Well, perhaps a story acceptable in physics nowadays does (although I need to be convinced) but this should not be enough to convince us that physics does get the *dramatis personae* of its causal stories right.

We are more confident about the ontology of physics than about the very general descriptions physics appears to try to give of it. We feel that Cartwright is right when she says '. . . the electron is not an entity of any particular theory . . . about which we have a large number of incomplete and conflicting theories . . .' (p. 92) and that an explanation which appeals to the existence of electrons 'makes no sense at all without the direct implication that there are electrons . . .' (p. 92). But I doubt whether an appeal to a fallacious mode of reasoning—inference to the best *X*—will help explain the feeling.

4 QUANTUM MECHANICS

I suspect quantum mechanics is the source of Cartwright's anti-realism. Her philosophy of quantum mechanics contains two important new ideas. First, that there is *no measurement problem* as such. Secondly, that all quantum probabilities are *transition probabilities.*

In quantum mechanics we find that the state-vector (for the individual system or pure ensemble) or the density operator (for the pure or mixed ensemble) evolves unitarily except when it doesn't, which is when a pure-state will be turned into a mixture. Most philosophers and physicists who have looked at the problem assume that it doesn't only during measurement. Cartwright claims that it doesn't at many other times as well. Reductions of the wave-packet occur in *state-preparations* (p. 174), in the Stanford Linear Accelerator (*SLAC*) all the time (pp. 172–4) and in decay and scattering processes (p. 179). The problem for the interpretation of quantum mechanics is, on Cartwright's view, not to account for the special features of measurement but to *characterise* the special conditions for the rather general non-unitary evolution of the state-vector or density operator we find in Nature.

Cartwright nowadays begins with the *fact* of non-unitary evolution—violation of Schrödinger's equation—a fact which in her old realist days she rejected. The Schrödinger equation predicts that a measurement on a quantum system will generally leave the measuring apparatus in a superposition of states. But we know that measuring apparatuses—chunks of macroscopic matter—are in determinate states. Many measurements made with a collection of measuring apparatuses (each used once) will result in a collection of apparatuses in a variety of different final states each of which is determinate. The result will be described as a *mixture* and not a superposition. In 1957 Feyerabend suggested that if superposition and mixture were *statistically equivalent* for macroscopic systems then one would have no reason to prefer the mixture description to the superposition description, and so one could preserve Schrödinger's equation. In 1962 *DLP* proved Feyerabend's conjecture (for macroscopic systems, systems with many degrees of freedom). It seemed possible to be realist about Schrödinger's equation and yet tolerate the use of mixtures for the description of measurement results. This was Cartwright's position of 1975 (p. 165).

But the statistical equivalence of superposition and mixture does not guarantee the behaviour of individuals in the ensemble will be equivalent.

Take Schrödinger's cat for example. Suppose that after one hour half the cats are dead, so that after two hours three-quarters are. The superposition description allows some of the cats dead after one hour to be alive after two hours. Only the proportions of dead and alive cats are the same in mixture and superposition. But 'macroscopic systems have orderly histories' (p. 171) and so the superposition description along with Schrödinger's equation has to go.

Having been liberated from this adherence to Schrödinger's equation Cartwright now goes much further. The Schrödinger equation is violated all over the place. Thus there is no need to appeal to the 'consciousness of the observer' to explain non-unitarity in measurement. The problem of non-unitarity is more serious as physics, but less serious as philosophy.

Why think that reduction of the wave-packet is widespread in physics? Cartwright cites the case of *SLAC*. In *SLAC* electrons surf down the accelerator on an electromagnetic wave, approaching the speed of light after a couple of miles (in our earthbound frame). But to surf they need to be localised. Yet as soon as a free electron (or indeed a surfing one) is localised it (or rather its state-vector in configuration space) will at once spread (almost everywhere) through the whole universe. To sustain our surfing picture the electron has to be continuously localised or 'collapsed'.

It may be that this process could be described, not as I have described it, but as falling under the Schrödinger equation. The description would be enormously complicated, no one would attempt it, and Cartwright seems to be saying that in fact it is impossible (p. 173). We *must* assume continuous collapse. I wonder. There is no very convincing argument, as there is against the DLP theory of measurement, that the reduction *must* be taking place, unless it is that the practice of physicists takes it to be. (And the philosophy of physics is, after all, the philosophy of *real* physics.)

The second of Cartwright's claims is that quantum probabilities are transition probabilities. This implies a drastic limitation on the interpretation of quantum mechanics: '[w]here no transitions occur, ψ must remain uninterpreted, or have only a subjunctive interpretation: if the system were subject to a perturbing potential of the right sort, the probability for a transition into state . . . would be . . .' (p. 186). In fact this seems circular, we explain what a state is (subjunctively) in terms of transitions into other *states*.

The argument begins with the two-slit experiment. This shows that quantum systems (photons say) must be treated sometimes as waves, sometimes as particles. In probabilistic terms the puzzle is this. Why does the equation

$$Prob[R \mid A \vee B] = (1/2)(Prob[R \mid A] + Prob[R \mid B]) \; [**]$$

fail? R is the proposition that the photon hits the screen at r, and A and B are the propositions that it goes exclusively through either slit 1 or slit 2, and $Prob[X \mid Y]$ is a conditional probability.

Cartwright is right to reject Putnam's naive quantum logical resolution (which he does too) which blocks the derivation of the equation by refusing the use of the distributive law used in deriving it. In fact, the distributive law does not fail, and if it did you would still have the paradox. But this does not dispose of all quantum logical interpretations of quantum mechanics, and any such will assign probabilities to propositions asserting that dynamical variables have values in ranges.

I should say that Putnam's most important error (apart from his quantum logical realism) lay in trying to combine quantum logic with classical probability with its unproblematic conditional probability. But can there be a quantum logical conditional probability? It so happens that all the conditional probabilities in [**] are zero if defined classically. Consider then this question. Can there be a viable quantum conditional probability? By viable I mean such as to be

(i) consistent with classical conditional probability when X and Y are compatible, but
(ii) not necessarily zero otherwise (even when X and Y are totally incompatible—incompatible for all states).

Put another way, can we define a quantum conditional probability $QProb[X \mid Y]$ such that for one lattice polynomial $p(X, Y)$

$$QProb[X \mid Y] = Qprob[p(X, Y)]/Qprob[Y]$$

where $QProb[X \mid Y]$ is a quantum conditional probability satisfying (i) and (ii), and $Qprob[X]$ is a quantum probability? The answer is that we can. Let $Com(X, Y)$ abbreviate 'X and Y are incompatible'. If $K(X, Y)$ is the polynomial

$$(X \wedge Y) \vee (X \wedge Y^{\perp})$$

we have in quantum logic

$$Com(X, Y) \quad \text{iff} \quad X = (X \wedge Y) \vee (X \wedge Y^{\perp}).$$

So let $p(X, Y) = (X \wedge Y) \vee (Y \wedge (K(Y, X))^{\perp})$. Then $QProb$ satisfies (i) and (ii).

But if one further demands the very reasonable

(iii) $QProb[X \wedge Y \mid Y] = QProb[X \mid Y]$

then it follows that there can be no non-zero quantum conditional probability for complementary (totally incompatible) pairs of propositions (like the pairs A and R, and B and R in the two-slit experiment).

Putnam's resolution of the two-slit experiment should have been to assert that the paradox cannot be formulated at all. Cartwright appears to take the level-headed line that quantum logic is all a mistake. But her solution to the paradox of the two-slit experiment is to restrict probability to transition probabilities, so that [**] cannot make sense. Ironically her solution is very similar to the one Putnam should have proposed.

Professor Cartwright's achievement is not that what she says on each or any issue is correct, but rather that her work will define the field for the next few years. And one cannot ask for more than that from any philosopher. One finds in her book a much-needed shift away from the abstract view of physics one finds in so much of the recent philosophical literature, and towards the nuts and bolts of real-live physics. Cartwright does not put much faith in the abstractions of mathematical physics, of which the Maxwellian distribution function may be taken as an example. There is a Wittgensteinian flavour to the book—something of 'Don't think, look!'—which sometimes affects her prose: [W]hat is the distribution function for the molecules in this room? Or the value of the electric field vector in the region just at the tip of my pencil? These questions are queer. They are queer because they are questions with no answers' (p. 156). The effect of Cartwright's book will be to make the philosophy of physics more sensible but also, for us philosophers, that much more complicated, that much harder.

PETER GIBBINS
Oxford University

REFERENCES

CREARY, L. [1981]: 'Causal Explanation and the Reality of Natural Component Forces', *Pacific Philosophical Quarterly*, **62**, pp. 148–57.
DANERI-LOINGER-PROSPERI [1962]: 'The Quantum Theory of Measurement and Ergodicity Conditions', *Nuovo Cimento*, **44B**, pp. 297–319.
HACKING, I. [1983]: *Representing and Intervening*. Cambridge University Press.
MILL, J. S. [1868]: *A System of Logic: Vol. I*. Seventh edition. Longmans.
RUSSELL, B. [1937]: *The Principles of Mathematics*. Allen and Unwin.

Reviews

BLACKWELL, K., BRINK, A., GRIFFIN, N., REMPEL, R. A. and SLATER, J. G. (*eds.*) [1983]: *Cambridge Essays 1888—1899* by Bertrand Russell. (Volume I of *The Collected Papers of Bertrand Russell.*) Allen and Unwin. xxxiv + 554 pp.

In my opinion, which differs from that of some critics in the weeklies, Russell is a sufficiently important philosopher to make it worth publishing his collected writings. But accepting this, one must also accept that a collection of twenty-eight volumes will have a Volume I which, since the collection is chronological, will be of less interest. This volume fully lives up to that. It consists of seven parts. The first comprises some Greek exercises and essays written at his crammer's, whilst the second contains personal writings of little interest. In the third part are to be found papers read to the Apostles, and the fourth and fifth reprint his fourth-year essays on epistemology, history of philosophy and ethics.

Papers of more philosophical interest begin in Part 6, with essays on the foundations of geometry, particularly four extracts from a notebook of 1895 with drafts of sections, later rewritten, of the fellowship dissertation, his paper *The Logic of Geometry* reprinted from *Mind* of 1896 and *The A Priori in Geometry* from the Aristotelian Society 1896. Since the fellowship dissertation (now lost) was submitted in 1895 and the published version came out in 1897, these papers show something of his changing ideas then. But what is shown is not, alas, of much significance, consisting mainly, as it does, in his curious transfer of his transcendental deduction from a deduction of general metrical geometry in 1896 to projective geometry in 1897. The seventh part of the book is on economics, mostly on German Social Democracy.

The most interest in this volume lies in the appendices and the critical apparatus. The list of books in Appendix II, *What Shall I Read*, gives useful information to the student of Russell. For example, it is now clear that already in December 1893 he was reading Erdmann's *Die Axiome der Geometrie*, and, two months later, Riemann's *Hypothesen welche der Geometrie zu Grunde liegen* although we don't have evidence of his being advised to write on 'Metageometry' for his fellowship until after the Moral Sciences tripos in May 1894. Or, again, in August 1894 he was reading Bradley's *Appearance and Reality* which is useful to know, since it gives an origin for some of Russell's ideas on the nature of the proposition. The appendices, including their author-index, take up 38 pages, the notes on the material 72, and the textual notes of a more technical paleographical nature 56. The rather heavyweight scholarship looks silly in this volume, but if it is a token of what is to come, it will be valuable in the later ones. Nothing is so trivial as to escape the interest of the editors (though there is an error on p. 403, with the note 102:39 'When Sigmund refuses Valhalla; an incident in Act I . . .' should, of course, read 'Act II'). The textual notes on the geometry papers mentioned above contain a useful table comparing their various sections with the corresponding ones in the published dissertation.

A group of philosophical papers on the philosophy of mathematics which are before 1899 are held over till Volume II.

C. W. KILMISTER
King's College London

REFERENCES

RUSSEL, B. A. W. [1896]: *Mind, n.s.* 5, pp. 1—23.
RUSSEL, B. A. W. [1896]: *Proc. Aristotelian Society*, o.s. 3 (no. 2), pp. 97—112.
RUSSELL, B. A. W. *An Essay on the Foundations of Geometry*, Cambridge.

HUSSEY, EDWARD (*ed.* and *tr.*) [1983]: *Aristotle's Physics Books III and IV*. Clarendon Press. xlix+226 pp. Hardback £13.50. Paperback £6.95.

Books III and IV of Aristotle's *Physics* form a connected series of inquiries into the topics of change, infinity, place, the void, and time. For sheer practice in hard fundamental thinking without benefit of formal apparatus, this material is a gift to doers of philosophy. It also sheds brilliant light on the concepts in question. The serious need for a good English translation-cum-commentary has now been more than met by Edward Hussey's edition. Hussey has prepared it with meticulous attention to the interests of readers at every level; the result is a fine work of scholarship and an excellent philosophical guide.

Twentieth-century readers may feel a certain impatience when first faced with Aristotle's definition of change at the start of Book III, especially if they compare it with Russell's definition (*Principles of Mathematics*, §442): 'Change is the difference, in respect of truth and falsehood, between a proposition concerning an entity and a time T and a proposition concerning the same entity and a time T', provided that the propositions differ only by the fact that T occurs in the one where T' occurs in the other'. With this available, who needs Aristotle's dictum: 'Change is the actuality of that which potentially is, *qua* such'? Well, anyone, perhaps, who seeks a theoretical basis for the fact that if x heats up so as to be hotter than y, we do not think that y has thereby altered, although something new is true of it; also for the fact that we do not think that the concept 'hotter than y' has altered, even though it is true of it now, and was false before, that it applies to x. Russell's elegant nutshell provides no backing for these intuitions; not so Aristotle's definition, which interlocks with other concepts to generate the theorem: 'Nothing changes that is not contiguous with an agent of the change'. Whatever else we may think of this, it gives the result that x in the example is the only subject of change. But the interest of the Aristotelian notion goes beyond the fact of its solving that puzzle. The failure, in that respect, of the 'Cambridge' criterion would have seemed unimportant at a time when rigorous analysis meant, in effect, reduction to the terms of *Principia Mathematica*. In that atmosphere Aristotle's definition could only seem a museum piece, and a grotesque one. Perhaps we are now better placed to appreciate its subtlety and power. It presents change from being F to being G as not reducible to a series of states of the subject. The corresponding proposition entails propositions about these states, but the former is not a truth-function of the latter. This is because 'F' and 'G' here specify the 'from which' and 'to which'; the directionality implied by this pair of prepositions is a structural feature of the context '– changes from – to –'. Thus direction of change, in a given case, is specified without additional reference to instants of time labelled according to their order in the time-series. Consequently, the way is open (and Aristotle takes it) to define time by reference to change. But to secure this advantage it is not enough simply to refrain from reducing change to a set of states (which then have to be assigned an order *ab extra*, so to speak). For if instead a change is treated as itself a unitary state, how can time, essentially plural, be reasonably defined in terms of change? In any case, if we classify '– changes from F to G' as a logically simple predicate on a par with '–F' and '–G', we face a tiresome and possibly vicious regress, since if x does change$_1$ from F to G, how is it not also the case that it changes$_2$ from being in change$_1$ to G, and from being in change$_2$ to G, *etc.*? For us these are all first order properties, whereas Aristotle's definition expresses the insight that the change-predicate is of a different logical type. The nature of the difference is spelt out precisely in terms of the categories and the 'actual'/'potential' distinction. An immediate consequence is that only expressions like 'F' and 'G' can properly function as direction-specifiers; to assign that role to a pair either of whose members is of the 'changes from F to G' form would be a mistake on a level with treating a relation as a term along with its terms.

Aristotle's discussion of the infinite had at least one enduring result: the distinction between 'endless' as implying a whole than which nothing is greater, and as applied to an uncompletable series. More controversial is his other famous conclusion, that in neither sense does an infinity actually exist. His refutation of the possibility of an infinite physical magnitude is now only of historical interest; there are other, more exciting, arguments based on his operationalist conception of number. Number-words, according to this, function primarily as adjectives applied through counting to sets of perceptible items. A countless set is thus literally *without* number; the point is not the easy one that 'greatest number' is self-contradictory, but that numbers are essentially for counting the countable. So the infinity of even the natural number series consists only in the fact that however high we count, we might have counted higher. This 'might have', though, is not just a matter of our own strength to keep going; nor does Aristotle let it depend on some dubious assumption to the effect that for any actual *n*-membered plurality of, say, sheep, it is physically possible that there should exist a group of $n+1$ members. Rather, he models the number series on the series of fractions produced in a continuous magnitude by repeated division in the same proportion. The divisions are not 'there' already: to be, they have to be made, but since each creates, in effect, a new divisible magnitude, further division is *necessarily* always possible. However, this necessity does in the end rest on a physical assumption: that physical magnitudes *are* infinitely divisible. For Aristotle rejects the idea that pure geometric magnitudes are real, for the same anti-Platonist reasons as govern his operationalist view of numbers. So is the infinity of the natural numbers propped up by a falsifiable empirical theory? He, at any rate, is saved anxiety here by his assumption that the materials of physical substances are utterly homogeneous, without structure at any micro-level. In a sense this is an empirical position, but (partly for metaphysical reasons) he takes it as certain, and so is assured of infinite divisibility without yielding to Platonism.

His account of place and the void is narrow and archaic, being heavily influenced by the theory that the various kinds of 'simple bodies' have 'natural motions' to specifically different 'natural places'. Even so, for modern readers his intense focus on certain issues also helps sharpen awareness of those he ignores. Starting from the datum that bodies can replace each other in the same place, he aims to establish (1) that a place is not an occupiable extension with its own reality. This might have led to a straightforward notion of place as relative location; except that it is complicated by his desire to accommodate the ideas (2) that x's place is what x is *in*, in the primary sense of 'in', and (3) that the place of x, strictly speaking, exactly fits the dimensions of x. These, given (1), lead to the definition of place as the inner surface of the surrounding body. But can one body be said in any sense, let alone the primary, to be *in* a surface of another? One feels that were this someone else's theory Aristotle would be first to object that although x is (straightforwardly) in another body y in virtue of contact at the latter's inner surface, that is no reason for treating the *surface* as something which x is in. This obsession with the inner surface of the surround as providing the metaphysical mainstay of x's 'in-ness' may explain his failure to raise questions about other aspects of the relation: questions, *e.g.*, that occur when one considers that y, to surround x, must also extend away from that inner surface. Thus he does not ask what it is for bodies to *be* physically extended, nor what it is for them to share common space so that contact and replacement are possible. However, interest in universal space as the formal condition of the availability in principle of any location to any body of suitable size would have opened a dangerous road for Aristotle: one that leads in the opposite direction from the physics of 'natural places'.

No such inhibition hampers his approach to time, since there is no exact temporal analogue to the concept of 'natural place'. He defines time as a sort of number; part of the point seems to be that a time, like a number, is indifferent, as to its formal structure, to the various empirical items to which it applies. His discussion of time is often frustratingly obscure: not only is the presentation rather disjointed, but one senses a fundamental disparity between his way of articulating the topic and that which sets the framework of contemporary discussions. For instance, he is impervious to the distinction (in McTaggart's terminology) between *A*-series and *B*-series. He sees spatial order (a matter not discussed in the section on place) and the continuity of extension as the ground of temporal order and continuity, with the concept of motion playing a mediating role between the concepts of extension and time. But this concern with the space-like properties of time never leads him to bracket them off in contrast with the 'dynamic' *A*-series attributes. On the contrary, he actually expounds temporal 'before and after' in terms of 'the now' which is the same yet 'always different and different'.

Hussey's translation is exact and it reads well (though see below on a few oddities). His lucid commentary is alert and informative on a wide range of conceptual and historical matters, as on the logical detail of Aristotle's arguments. Hussey has made this the occasion for a searching study of Aristotle's finitist philosophy of mathematics. This material is divided between introduction, notes and a long appendix. Hussey hopes to publish a fuller treatment elsewhere; that will be worth waiting for, but the present discussion is already an important contribution to research. Of particular interest is his use of 'possible worlds' to provide finitist foundations for the arithmetic of the natural numbers; also his 'localised' interpretation of the Euclidean references to infinite lines. A second long appendix is given over to an illuminating discussion of Aristotelian dynamics. Hussey argues that the texts imply a distinction between the pure mathematics of moving bodies and the empirical science of mechanics. On this basis Aristotle's statements on dynamics turn out to be a good deal more coherent than usually supposed.

This book contains a six-page bibliography; a list of divergences from Ross's text (all of which I found reasonable except for the excision of 202b 26–8); a generous glossary of Greek terms; an index of names; and a highly organized index of subjects.

I noticed minor misprints at page xlv, line 4 (of Hussey's page); 8, line 11; 36, line 27 (a square bracket is missing); 104–5 at the top, where 'III' should be 'IV'; 116, line 22; 144, line 17, where '*Physics IV*' should be '*Physics VI*'. At 40, line 31, the second 'rare' should be 'hard'. On 97, axiom (a) needs brackets, and the formula two lines below is not perspicuous for the same reason. At 132, line 25, '215a 13–14' should be '216a' *etc.*, and '216' should be added to the reference on line 35. Also, the note here has 'force' whereas the main translation says 'strength'. On translation: I was puzzled by 'dimensions' (why not 'directions'?) at 11, line 4; and by 'bounded' at 21, line 24 (the context demands 'determined'); also by 'extension *between* body' (my italics) at 33, lines 2 and 18. The phrase 'six dimensions' on pages 13 and 20 is the exact translation, but for us it has a misleading ring.

SARAH WATERLOW
University of Edinburgh

LIGHTFOOT, DAVID [1982]: *The Language Lottery: Toward a Biology of Grammars*. The M.I.T. Press. $17.50 xii+224 pp. and JULIA, PERE [1983]: *Explanatory Models in Linguistics: A Behavioral Perspective*. Princeton University Press. xv+227 pp.

Linguists have traditionally treated languages as structured systems of significant signs or symbols whose properties as systems can be investigated, described and explained in abstraction from the ways and circumstances in which these systems are put to use in ordinary communication. Linguists, in other words, have rarely (and only marginally) undertaken investigations of the typical speech situation in its full complexity: they have rarely undertaken to explain why speakers use the particular forms of words which they do use in the particular circumstances in which they speak; they have (usually) chosen instead simply to try to specify the system of correlations between sounds and meanings which is exploited in the typical speech situation.

This traditional emphasis on system, as opposed to use, was codified and rationalized by Saussure, who drew a distinction between *langue*, the system of sound-meaning correlations, and *parole*, the exploitation of this system in ordinary language use, and who argued that *langue* is systematically structured, whereas *parole* involves the free and creative activity of speakers and is therefore unsuitable as an object of scientific investigation.

More recently, Chomsky has accepted and extended Saussure's insights. Corresponding, roughly, to Saussure's *langue/parole* distinction is Chomsky's distinction between linguistic competence, the mentally embodied systematically structured set of correlations between sounds and meanings, and linguistic performance, the exploitation of this system in ordinary language use. Like Saussure, Chomsky has also urged that direct investigations of language use (*parole* or performance) are condemned to vacuity in view of the fact that ordinary language use appears not to be under the systematic control of external, independently identifiable stimuli: we appear to be unable to establish reliable nonvacuous correlations between characteristics of speech situations and characteristics of the linguistic forms which are produced in those situations.

Linguists working in the Chomskian tradition have thus been primarily concerned to investigate linguistic system, rather than language use.

Lightfoot endorses this standard investigative strategy and provides, for the general reader, a vivid and accessible account of some important subsidiary methodological principles. Julia, on the other hand, rejects this entire point of view and, in particular, questions the acceptability of Chomsky's initial abstraction away from language use to linguistic system. According to Julia (p. 15), 'the basic question [*pace* Chomsky] remains the same: Why does verbal behavior occur when it does and why does it have the form it has on every particular occasion?'

If Julia could establish, *contra* Chomsky, that this basic question does in fact have a non-vacuous answer, then his work would, of course, be of the greatest importance. But Julia does not in fact establish this. Nor, indeed, does he provide anything more than the sketchiest account of how this question might be fruitfully addressed. He does claim (p. 13) that this question can and should be addressed in terms of a Skinnerian account of language use which 'places the burden of explanation outside the organism, in the controlling relations responsible for the rise and maintenance of verbal behavior'. But he provides no compelling (or original) evidence or arguments to support this general claim in the face of objections to this behaviouristic program of explanation which Chomsky and others have repeatedly made since the late 1950s. More to the point, perhaps, Julia provides no concrete examples of adequate explanations of language use of this behaviouristic character. Not least among the other defects of Julia's book is his rejection, in the absence of convincing or original arguments, of some of the major methodological principles which are routinely employed in linguistic research and which Lightfoot so lucidly outlines, illustrates and defends.

Lightfoot, then, provides particularly clear and forceful discussions of three of the most important features of the methodology of Chomskian linguistics. These three closely related features are the principles of modularity and idealization and the requirement of independent testability.

The modularity principle is, in effect, embodied in a heuristic principle, according to which complex phenomena are to be explained as composite effects of the interaction of distinct systems of causes, rather than being treated as simple and direct effects of some simple system of causes. (Julia says (p. 18) of this principle that we should reject it and instead 'face verbal activity in all its functional complexity'.) As Lightfoot demonstrates (pp. 42 *ff.*), this general methodological principle has a natural application in linguistics when we attempt to account for the intuitive judgments which language users make about the sentences of their languages. For instance, speakers of English are likely to find sentences such as THAT THAT THAT THE MOON IS BRIGHT IS OBVIOUS DISTURBED ME SURPRISED HARRY intuitively unacceptable. On a non-modular approach, we would treat this fact as one which ought to be captured by any adequate theoretical account of English and would therefore reject any grammar of English, whatever its other merits, which 'generates' this sentence. On a modular approach, however, we might not be forced to do this. Rather, we might treat this common intuition as a composite effect of the interaction of two distinct mental systems—a grammar which *does* 'generate' this sentence and a general perceptual system which is unable easily to assign linguistically significant structure to this sentence, despite the fact that it is 'generated' by the grammar.

The idealization principle is likewise embodied in a heuristic directive, according to which complex phenomena are to be explained in terms of ideal types which are constructed by abstracting away from explanatorily irrelevant variations from a norm, rather than being treated in their full complexity. (Julia says (p. 128) of this principle that 'nothing justifies shying away from the complexities of real data'.) As Lightfoot demonstrates (pp. 90 *ff.*), this general methodological principle has a natural application in linguistics when we attempt to produce an account of linguistic system in the face of widespread variations in linguistic competence among the members of a given speech community. For instance, judgments of the intuitive acceptability of sentences are known to vary from speaker to speaker in many linguistic communities. On a non-idealizing approach, we would treat such variation as a fact which ought to be captured by any adequate account of the language of a community, and would therefore reject any grammar which failed to account for this variation. On an idealizing account, however, we might instead simply abstract away from such variation as had been detected and direct our attention to those features of linguistic system which are common to the competences of the members of such a community.

(There is, of course, an important relation between the principles of modularity and idealization which needs to be mentioned. It is this: modularity may well (and usually should) be invoked in support of idealization. When we construct an ideal type by abstracting away from variation in a concrete system, the claim that the ideal type explains the behaviour of the system in its full complexity will clearly be strengthened if we can, in addition, show that the variation which we have 'factored out' can be accounted for in terms of an interaction between factors identified by the ideal type and factors associated with another, 'interfering' system, which we might also characterize in terms of an ideal type.)

The requirement of independent testability is, of course, familiar from general methodological theory. But as Lightfoot shows (pp. 45 *ff.*), in effect, this requirement acquires special significance in the linguistic context, where modularity and idealization principles are so widely employed. This is so, in particular, because both these principles are subject to abuse which only this requirement can prevent. For instance, we could easily enough insulate nearly any grammar even from massive refutation by incompatible intuitive linguistic judgments simply by invoking the modularity assumption and attributing this incompatibility to an 'interfering' system. The requirement of independent testability militates against such a methodologically untoward practice however. It does so by requiring ***independent*** confirmation of the claim implicit in such a modular hypothesis: that the 'interfering' system is responsible for the incompatibility between the consequences of the grammatical theory and the intuitive judgments at issue.

These, then, are some of the basic methodological tools employed by linguists working in the Chomskian tradition. As I have said, Lightfoot provides a clear and concretely illustrated account of these tools and of the ways in which they are put to use. Julia, on the other hand, rejects the project in which they are wielded, though he also seems, inconsistently, to agree with Chomsky and Lightfoot, that (p. 134) 'science... proceeds in steps, approaching complex phenomena through the analysis of their contributing simpler parts'.

FRED D'AGOSTINO
Research School of Social Sciences
Australian National University

MAZE, J. R. [1983]: *The Meaning of Behaviour*. George Allen and Unwin. vi+186 pp.

Rage is a natural human passion, and voiding it cleanly is fun; it is, I suspect, an engine which drives some, including apparently some psychologists, to philosophical criticism. Another man's rage can entertain and, on occasion, move us. But unrelieved bombast soon draws only contempt, not sympathetic rage; it is too little cooled by intelligence in spurting from its infantile sources. Wit stirs the anger restrained by intelligence without breaking the restraints; thus, whatever our practice, we prefer the idea of Horace to the idea of Juvenal.

The pythoness at Delphi sat on a tripod to deliver her oracles. Maze's tripod is the conventional one amongst us, Science, and its three feet seem to be empiricism, determinism and central state materialism. When behaviour is decoded according to this legend, its meaning seems to be this: There are innate in the brain certain drive centres. We do not know what these are, but they are best specified by their bodily sources and probably include sexuality, hunger, thirst, respiration, avoiding activation of the pain nerves, and temperature control. They are natural engines; certain keys fire them up, and as they chug along, they cause the organism to motor about the world. In causal transactions between the nervous system and the rest of the world, traces are left in that system which can somehow hook up with its native engines, either fore or aft; in this way, both the conditions under which the engines fire up and their subsequent arabesques can become baroquely elaborated. The result is human life. In a sense, this rendering follows our understanding of actions in terms of desires and beliefs. The engines are the instincts which, elaborated, we misconceive as our desires, while the traces left by the world on our nervous systems are our beliefs.

And that, plus a few references to feedback, is that. Maze's sketch of mechanisms in the central nervous system is certainly a sketch of deterministic causation realised in matter; and if we could trace out these long reflex arcs, watching them pulse would be as empirical as observation gets. So Maze appeases Science. But his sketch is far and away the prevailing orthodoxy among contemporary philosophers, however opinion may sort itself out among psychologists. So how does his engine elaborate into a book; and anyhow, what justifies faith in its sketch of a blueprint? That is where the rage comes in.

First and foremost, his is an offering to Science, and while Her metropolitans may offer Her substance, the more peripheral Her acolytes, the more they are reduced to propitiating Her with invective against their superstitious neighbours. But then, what other service is there? Isn't anything other than Science mere superstition? In the large, Maze's is a what-else argument: the alternatives to his account are thus and such; but thus is rubbish; and such is fatuous; so what else is hard sense but his account? Once upon a time Hilary Putnam used to teach philosophy students to suspect what-else arguments; they are no better than the range of alternatives they suppose, and that range merits no more trust than the philosophical imagination sporting in a wilderness not yet pruned to perspicuity by Science, by an explanation of why those and only those are the alternatives.

Maze puts no trust in intentionality. For the mentality of a mental state to be constituted by its intentionality is for it to be constituted by 'its' relations to other things which, alarmingly, may not exist. But nothing could be constituted wholly by its relations to other things. Each thing must have its own intrinsic nature; relations arise through transactions between things by virtue of their natures. So, at any rate, Maze seems to argue.

At this point I began to wonder whether questions arise about the preferable sort of mind-body identity theory. Grant that we describe beliefs by their (intentional, analogous to semantic) relations with their subject matter; this is an epistemic point, and may be no more than that. Turning to metaphysics, any identity theorist presumably holds that for each belief, there is something or other in the central nervous system with which the belief is identical; and that for whatever is identical with the belief, there is a nature both share. A type-type identity theorist could also hold that for each belief there is a nature which it shares with whatever in the nervous system is identical with the belief; such shared natures found psycho-physical laws and make possible a natural science of interpretation. Token-token identity theorists who deny that there are any genuine psycho-physical laws presumably deny the type-type theorist's *AEA* thesis, and may owe us an apology for interpretation. It is very difficult to tell whether Maze takes a position in this dispute, but his critique of representation suggest that he might take the token-token theorist's side. Indeed, despite occasional pronouncements otherwise, much of the time he seems to mean that ultimately there will be no mention of beliefs in explanations of actions; neurophysiology will explain *away* action and belief. Here it might have been interesting to have had a psychologist's critique of Davidson's argument in 'Mental Events' from rationality against psycho-physical laws, or of Kaplan's causal theory in 'Quantifying In' of the rapport between a person with a *de re* belief and the subject matter of his belief; subtle and complex thought invite subtle and complex responses. Instead, Maze tells us that 'dual-aspect' conceptions of mind-body relations, according to which some nervous processes have 'mental properties' which cause them to work differently from nervous processes that lack them, do not deserve serious consideration. 'It is mere verbalism to speak of "mental properties of nervous processes" if by that is meant localised, embodied, intrinsic properties; any imaginable properties of nerve cells and events in them are plain physical ones' (p. 57). While Davidson argued from the incongruity between holistic attributions of rational propositional attitudes and one-by-one attributions of physical states to the impossibility of psycho-physical laws, Maze preaches a sermon against the 'peculiar aura of sanctity attaching to the word "holistic" ' (p. 34); he makes no serious attempt to articulate the virtues and vices implicit in the notion of holism; he betrays no sensitivity to the analogues of inverted spectrum examples, like putting this in your mouth either because you are hungry and think this is food or because you are suicidal and think this is poison, which might prompt considerations of holism. (But he does invoke similar considerations in chapter six to claim that drives must be specified by their bodily sources.) Maze has no patience with subtlety, largely, I think, because he enjoys venting impatience.

He fulminates against desire, intention, purpose, choice and agents not simply on the ground that they evaporate through being constituted by their relations to other things. All talk of the appetites and the will is incoherent because it involves the idea of an agent who has several really possible options open, who chooses one and who applies himself to doing it. But because of determinism, there is never any real choice about what to do; and self-starters are unintelligible because there would be nothing prior in terms of which they could be explained. The only way not to be 'pre-scientific, pseudo-explanatory and inescapably obscurantist' (p. 36) is to suppose in people's bodies nothing but mechanisms which, turned on through efficient causality, move those bodies. There is, I think, some sort of good sense here, but my willingness to make the mechanist supposition weakens the louder it is shouted at me. Indeed, the shout causes in me impious worries about whether, and if so why, people must ultimately be intelligible.

When he turns on his peers, he goes to town. He ridicules behaviourists for refusing to look inside the black box, and on Chomsky's ground that there is no general account of what reinforces save what pleases. He ridicules the phenomenologists and the existentialists for their 'inability to generate any useful or intelligible psychological theory' (p. 126); here, he rounds up the usual suspects, like the transcendental ego, freedom and authenticity. I am not sure that there is any more argument here than Sartre calling determinists like Maze stinkers, or whatever Sartre's technical term was. I wish a gifted analytic philosopher would write a sensitive and sympathetic critique of phenomenology and existentialism; I am tired of my youthful prejudices.

Maze admires Freud; indeed, Maze uses his own views to re-work Freud's. Parents so arrange a child's world that satiating some instinctual drives frustrates satiating others. This eventually splits the drives into two groups, one the ego and the other the id; the former are those of which the parents approve and the latter are those subject to parental disapproval. As the child grows up, all this is repressed and replaced by the delusion of virtue whose economic function is that 'if he retained in consciousness the realisation that he had given in to the arbitrary demands of others, that would be humiliating and painful' (p. 171). Of course, none of this can be explaining action by giving its motives. Good art and science, unlike morality, are expressions of the repressed.

An impressive idea of Freud's is his programme to explain the detail of dreams, neurotic symptoms and mistakes like slips of the tongue on the hypothesis that they are actions whose detail is accounted for by the detailed content of the beliefs and desires invoked to explain them; no other psychology explains the detail of these phenomena. If the propositional attitudes lack natures, and thus details linked by psycho-physical laws to aspects of neurological mechanisms, will Maze have to give up the Freudian enterprise of interpretation? Or does Maze want from psychoanalysis only a device with which to *épater les bourgeois*?

The distinctions between the disciplines and departments are largely illusions imposed on the life of the mind by university administrators mostly for budgetary reasons. Hence, the idea of a psychologist conversant with philosophy writing about action is attractive. Anything which reduces ignorant contempt between disciplines is good, anything, that is, except perhaps undiluted expressions of contempt.

W. D. HART
University College London

Recent Publications

ADKINS, Bruce M. (*ed.*) [1983]: *Man and Technology.* Cambridge Information and Research Services Ltd. Paperback £15.00. 0-905332-30-X. Pp. xx + 300.

ANDERSSON, Gunnar (*ed.*) [1984]: *Rationality in Science and Politics.* D. Reidel, Dordrecht. Hardback Dfl.130.00. $49.50. 90-277-1575-0. Pp. 307.

ARMSTRONG, D. M. and MALCOLM, Norman [1984]: *Consciousness & Causality. A Debate on the Nature of Mind.* Basil Blackwell, Oxford. Hardback £17.50. 0-631-13212-0. Paperback £6.50. 0-631-13433-6. Pp. 222.

ARONSON, Jerrold L. [1984]: *A Realist Philosophy of Science.* Macmillan Press. Hardback £25.00. 0-333-35329-3. Pp. x + 278.

ASMIS, Elizabeth [1984]: *Epicurus' Scientific Method.* Cornell University Press, Ithaca and London. Hardback $49.50. 0-8014-1465-2. Pp. 385.

BENENSON, F. C. [1984]: *Probability, Objectivity and Evidence.* Routledge & Kegan Paul, Oxon. Hardback £19.95. 0-7100-9598-8. Pp. ix + 284.

BERGGREN, W. A. and VAN COUVERING, J. A. (*eds.*) [1984]: *Catastrophes and Earth History. The New Uniformitarianism.* Princeton University Press, N.J. Hardback £60.10. 0-691-08328-2. Paperback £13.90. 0-691-08329-0. Pp. xii + 464.

BEVILAQUA, F. and KENNEDY, P. J. (*eds.*) [1983]: *Using History of Physics in Innovatory Physics Education.* (*International Commission on Physics Education, Pavia, Italy, 1983*). Universita di Pavia. Paperback $10.00. 0-8507510-2-2. Pp. 329.

BOAKES, Robert [1984]: *From Darwin to Behaviourism. Psychology and the Minds of Animals.* Cambridge University Press. Hardback £35.00. $69.50. 0-521-23512-X. Paperback £15.00. $19.95. 0-521-28012-5. Pp. xiv + 279.

BOGDAN, Radu J. (*ed.*) [1984]: *D. M. Armstrong.* (*'Profiles' series*). D. Reidel. Hardback Dfl.145.00. $55.00. 90-277-1657-9. Pp. ix + 304.

BRACKEN, Harry M. [1984]: *Mind and Language. Essays on Descartes and Chomsky.* Foris Publications, The Netherlands/U.S.A. Paperback Dfl.38.50. $15.40. 90-6765-020-X. Pp. xiv + 154.

BROWN, Robert [1984]: *The Nature of Social Laws. Machiavelli to Mill.* Cambridge University Press. Hardback £22.50. $39.50. 0-521-25782-4. Pp. ix + 270.

BUTTS, Robert E. [1984]: *Kant and the Double Government Methodology.* D. Reidel. Hardback Dfl.115.00. $43.00. 90-277-1760-5. Pp. xvi + 339.

CAMPBELL, Jeremy [1984]: *Grammatical Man. Information, Entropy, Language and Life.* Pelican Books, England. Paperback £3.95. 0-14-022504-8. Pp. 319.

COHEN, Robert S. and WARTOFSKY, M. W. (*eds.*) [1984]: *Hegel and the Sciences.* D. Reidel. Hardback Dfl.160.00. $61.50. 90-277-0726-X. Pp. ix + 374.

COHEN, Robert S. and WARTOFSKY, M. W. (*eds.*) [1984]: *Physical Sciences and History of Physics.* D. Reidel. Hardback Dfl.105.00. $39.50. 90-277-1615-3. Pp. ix + 259.

EVANS, Brian and WAITES, Bernard [1981]: *I.Q. and Mental Testing. An Unnatural Science and its Social History.* Macmillan Press. Hardback £15.00. 0-333-25648-4. Paperback £5.95. 0-333-25649-2. Pp. x + 228.

FRIED, Yehuda and AGASSI, Joseph [1983]: *Psychiatry as Medicine. Contemporary Psychotherapies.* Martinus Nijhoff, The Netherlands. Hardback Dfl.80.00. £20.25. 90-247-2837-1. Pp. ix + 192.

GALTON, Antony [1984]: *The Logic of Aspect. An Axiomatic Approach.* Clarendon Press, Oxford. Hardback £18.00. 0-19-824430-4. Pp. viii + 160.

GENSLER, Harry J. [1984]: *Godel's Theorem Simplified.* University Press of America. Hardback $12.75. 0-8191-3868-1. Paperback $6.75. 0-8191-3869-X. Pp. 83.

GERHART, Mary and RUSSELL, Allan M. [1984]: *Metaphoric Process. The Creation of Scientific and Religious Understanding.* Texas Christian University Press, U.S.A. Hardback $16.95. 0-912646-82-9. Paperback $10.95. 0-912646-86-1. Pp. xvi + 217.

GIRALT, Eudaldo Forment [1984]: *Introducción a la metafisica del Lenguaje.* Universitat de Barcelona. Paperback 84-7528-121-4. Pp. 202.

GIRALT, Eudaldo Forment [1984]: *Fenomenologia Descriptive del Lenguaje.* Promociones Publicaciones Universitarias, Barcelona. Paperback 84-86-130-33-6. Pp. 352.

GOODMAN, Nelson [1984]: *Of Mind and Other Matters.* Harvard University Press. Hardback £14.90. 0-674-63125-0. Pp. 210.

GOULD, Stephen Jay [1984]: *The Mismeasure of Man.* Pelican Books, England. Paperback £2.95. 0-14-02-2501-3. Pp. 352.

GREGORY, Richard L. [1984]: *Mind in Science. A History of Explanations in Psychology and Physics.* Peregrine Books, England. Paperback £7.95. 0-14-055-168-9. Pp. xvii + 641.

HABERMAS, Jürgen [1984]: *The Theory of Communicative Action. Vol. One: Reason and the Rationalization of Society* (translated by Thomas McCarthy). Heinemann Educational Books, London. Hardback £25.00. 0-435-82391-4. Paperback 0-435-82392-2. Pp. xlii + 465.

HARE, J. D. [1983]: *Starting from Scratch.* J. D. Hare, Butztown, U.S.A. Paperback £3.00. Pp. 25.

HARRÉ, Rom and REYNOLDS, Vernon (*eds.*) [1984]: *The Meaning of*

Primate Signals. Cambridge University Press. Hardback £25.00. $39.50. 0-521-25944-4. Pp. xi + 257.

HENDRY, John [1984]: *The Creation of Quantum Mechanics and the Bohr-Pauli Dialogue*. D. Reidel. Hardback Dfl.95.00. $34.50. 90-277-1648-X. Pp. xi + 177.

HÜBNER, Kurt [1983]: *Critique of Scientific Reason*. (Translated by Paul R. Dixon Jr. and Hollis M. Dixon). The University of Chicago Press. Hardback £22.00. 0-226-35708-2. Paperback 0-226-35709-0. Pp. xii + 283.

JAMES, William [1983]: *The Principles of Psychology*. Harvard University Press. Paperback £14.85. 0-674-70625-0. Pp. xxi + 1,302.

JARDINE, N. [1984]: *The Birth of History and Philosophy of Science*. Cambridge University Press. Hardback £32.50. 0-521-25226-1. Pp. ix + 301.

JAVUREK, Z., URSUL, A. D. and ZEMAN, J. (*eds.*) [1984]: *Integration of Science and the Systems Approach*. Elsevier Science Publishing Co. in co-edition with Academia, Prague. Hardback Dfl. 145.00. $55.75. 0-444-99633-8. Pp. 258.

JAY, Martin [1984]: *Adorno*. Fontana Paperbacks, London. Paperback £2.50. 0-00-636287-7. Pp. 192.

KYBURG Jr., Henry E. [1984]: *Theory and Measurement*. Cambridge University Press. Hardback £22.50. 0-521-24878-7. Pp. viii + 273.

LAUDAN, Rachel [1984]: *The Nature of Technological Knowledge. Are Models of Scientific Change Relevant?* D. Reidel. Hardback Dfl.65.00. $24.95. 90-277-1716-8. Pp. vii + 145.

LEVINSON, Paul (*ed*). [1982]: *In Pursuit of Truth. Essays on the Philosophy of Karl Popper on the Occasion of his 80th Birthday*. Humanities Press, N.J. Hardback $25.00. 0-391-02609-7. Pp. xii + 337.

LINGUITI, Gennar Luigi [1984]: *Leibniz e La Scoperta Del Mondo Microscopico Della Vita*. Maria Pacini Fazzi Editore, Lucca, Italy. Paperback. Pp. 135.

LONGAIR, M. S. [1984]: *Theoretical Concepts in Physics*. Cambridge University Press. Hardback £25.00. $49.50. 0-521-25550-3. Paperback £8.95. $14.95. 0-521-27553-9. Pp. xiii + 366.

LOPREATO, Joseph [1984]: *Human Nature & Biocultural Evolution*. George Allen & Unwin. Hardback £20.00. 0-04-573017-2. Pp. xiv + 400.

MACDONALD ROSS, George [1984]: *Leibniz*. Oxford University Press. Hardback £7.95. 0-19-287621-X. Paperback £1.95. 0-19-287620-1. Pp. 121.

MARGOLIS, Joseph [1984]: *Culture and Cultural Entities: Towards a New Unity of Science*. D. Reidel. Hardback Dfl.80.00. $34.95. 90-277-1574-2. Pp. xiii + 170.

MARTIN, R. M. [1984]: *Mind, Modality, Meaning, and Method*. State University of New York Press. Hardback $44.50. 0-87395-721-0. Paperback $16.95. 0-87395-722-9. Pp. xiv + 225.

MAXWELL, Mary [1984]: *Human Evolution. A Philosophical Anthropology.* Croom Helm Ltd., Kent, England. Hardback £19.95. 0-7099-1792-9. Paperback £9.95. 0-7099-1794-5. Pp. 374.

MCGUINNESS, Brian (*ed.*) [1984]: *Wittgenstein and the Vienna Circle. Conversations Recorded by Friedrich Waismann.* Basil Blackwell, Oxford. Paperback £6.50. 0-631-13469-7. Pp. 266.

NOSSOL, Mgr. Alfons, *et al.* [1983]: *Collectanea Theologica.* Academia Theologiae Catholicae, Warsaw, Poland. Paperback. ISSN-0137-6985. Pp. 194.

PARSONS, Charles [1984]: *Mathematics in Philosophy: Selected Essays.* Cornell University Press, N.Y. Hardback $42.50. 0-8014-1471-7. Pp. 365.

PICKERING, Andrew [1984]: *Constructing Quarks. A Sociological History of Particle Physics.* Edinburgh University Press. Hardback £20.00. 0-85224-458-4. Pp. xi+468.

PRIGOGINE, I. and STENGERS, I. [1984]: *Order Out of Chaos. Man's New Dialogue with Nature.* William Heinemann, London. Hardback £9.95. 0-434-60395-3. Pp. xxvi+349.

RIEDL, Rupert [1984]: *Biology of Knowledge. The Evolutionary Basis of Reason.* John Wiley & Sons, Chichester, England. Hardback. 0-471-10309-8. Pp. ix+252.

ROSE, H. E. [1984]: *Subrecursion. Functions and Hierarchies.* Clarendon Press, Oxford. Hardback £22.50. 0-19-853189-3. Pp. xiii+189.

RUBINSTEIN, Robert A., LAUGHLIN, Charles D., Jr. and MCMANUS, John [1984]: *Science as Cognitive Process. Toward an Empirical Philosophy of Science.* University of Pennsylvania Press. Hardback £18.95. 8122-7911-5. Pp. xxv+186.

RUSSELL, Bertrand (EAMES, E. R. and BLACKWELL, K. *eds.*) [1984]: *Theory of Knowledge. The 1913 Manuscript.* George Allen & Unwin, London. Hardback £35.00. 0-04-920073-9. Pp. lv+258.

SCHNÄDELBACH, Herbert [1984]: *Philosophy in Germany 1831–1933.* (Translated by Eric Matthews). Cambridge University Press. Hardback £24.00. 0-521-22793-3. Paperback £7.50. 0-521-29646-3. Pp. x+265.

SEIELSTAD, George A. [1983]: *Cosmic Ecology: The View from the Outside In.* University of California Press. Hardback £19.20. $24.95. 0-520-04753-2. Pp. xi+169.

SHALLIS, Michael [1984]: *The Silicon Idol. The Micro Revolution and its Social Implications.* Oxford University Press. Hardback £8.95. 0-19-215877-5. Pp. vii+188.

SHAPERE, Dudley [1984]: *Reason and the Search for Knowledge.* D. Reidel. Hardback Dfl.155.00. $59.50. 90-277-1551-3. Pp. xlv+438.

SHOEMAKER, Sydney and SWINBURNE, Richard [1984]: *Personal Identity. The Dualist Theory.* Basil Blackwell, Oxford. Hardback £15.00. 0-631-13208-2. Paperback £5.50. 0-631-13432-8. Pp. 158.

SKALA, Heinz J., TERMINI, S. and TRILLAS, E. (*eds.*) [1984]: *Aspects of Vagueness*. D. Riedel. Hardback Dfl.105.00. $39.50. 90-277-1692-7. Pp. vii + 304.

SMITH, Joseph Wayne [1984]: *Reductionism and Cultural Being. A Philosophical Critique of Sociobiological Reductionism and Physicalist Scientific Unificationism.* Martinus Nijhoff, The Netherlands. Hardback Dfl.150.00. £38.25. 90-247-2884-3. Pp. 387.

SOBER, Elliott [1984]: *Conceptual Issues in Evolutionary Biology.* The M.I.T. Press, Cambridge, Mass. Hardback £36.00. 0-262-19220-9. Paperback £17.95. 0-262-69084-5. Pp. xiv + 725.

STICH, Stephen P. [1983]: *From Folk Psychology to Cognitive Science. The Case Against Belief.* The M.I.T. Press. Hardback £20.25. 0-262-19215-2. Pp. xii + 266.

STROUD, Barry [1984]: *The Significance of Philosophical Scepticism.* Clarendon Press, Oxford. Hardback £15.00. 0-19-824730-3. Paperback £6.95. 0-19-824761-3. Pp. xiv + 277.

TAROZZI, Gino (*ed.*) [1983]: *Gli Strumenti Nella Storia e Nella Filosofia della Scienza.* Istituto per i Beni Artistici Culturali Naturali della Regione Emilia-Romagna, Italy. Paperback. Pp. 272.

TOOMER, G. J. [1984]: *Ptolemy's Almagest. Translated and Annotated by G. J. Toomer.* Gerald Duckworth, England. Hardback £55.00. 0-7156-1588-2. Pp. ix + 693.

VAN DER MERWE, Alwyn (*ed.*) [1983]: *Old and New Questions in Physics, Cosmology, Philosophy, and Theoretical Biology. Essays in honor of Wolfgang Yourgrau.* Plenum Publishing Corporation. Hardback $95.00. 0-306-40962-3. Pp. xv + 920.

WESLEY, James Paul [1983]: *Causal Quantum Theory.* Benjamin Wesley. Paperback $24.00. 3-9800-942-0-0. Pp. xxiii + 405.

YOLTON, John W. [1984]: *Thinking Matter. Materialism in Eighteenth-Century Britain.* Basil Blackwell, Oxford. Hardback £19.50. 0-631-13335-6. Pp. xiv + 238.

ZUMBACH, Clark [1984]: *The Transcendent Science. Kant's Conception of Biological Methodology.* Martinus Nijhoff Publishers. Hardback Dfl.85.00. £21.75. 90-247-2904-1. Pp. xii + 165.

Philosophy and Sociology of Science

An Introduction

STEWART RICHARDS

Modern science is inseparable from the society that supports it. Yet now more than ever before the scientist is encouraged to specialize and narrow his field of vision. Stewart Richards argues that it is vital to reverse this trend, and that all students of science — natural and social — should look at the context, roots and structure of their disciplines.

220 pages, hardback **£17.50** (0 631 13411 5)
paperback **£5.95** (0 631 13414 X)

From Knowledge to Wisdom

A Revolution in the Aims and Methods of Science

NICHOLAS MAXWELL

In this wide-ranging yet closely argued book, the author argues forcefully the need for a comprehensive revolution in the aims and methods of science. He urges a new kind of enquiry which would have as its fundamental aim the enhancement of personal and social wisdom.

320 pages, **£19.50** (0 631 13602 9)

Rationality and Relativism

Edited by MARTIN HOLLIS and STEVEN LUKES

This book takes its inspiration from the debate encapsulated in *Rationality* (1970) edited by Bryan Wilson, but extends it into new fields and into current topics of discussion. Ernest Gellner, Martin Hollis, Robin Horton and Steven Lukes add to their earlier thoughts. They are joined by Barry Barnes, David Bloor, Jon Elster, Ian Hacking, W. Newton-Smith, Dan Sperber and Charles Taylor.

320 pages, hardback **£18.50** (0 631 12773 9)
paperback **£6.95** (0 631 13126 4)

Beyond Objectivism and Relativism

RICHARD J. BERNSTEIN

Philosophy is again questioning its goals and assumptions, a malaise signified by the growing opposition between objectivism and relativism, between rationality and irrationality, between realism and anti-realism. Contemporary thinking is becoming increasingly polarized and attempts to escape from this entire framework of analysis have all too frequently been assimilated to these standard oppositions. This major new book represents a courageous attempt to break out of this mould.

302 pages, hardback **£19.50** (0 631 13429 8)
paperback **£7.50** (0 631 13437 9)

Science and Moral Priority

The Merging of Mind, Brain and Values

ROGER SPERRY

'I consider Sperry's book to be of overwhelming significance.' M.E. Grenander, *Methodology and Science*

'Remarkable book whose philosophical repercussions are of quite exceptional importance.' Bernard Dixon, *Books and Bookmen*

152 pages, hardback **£15.00** (0 631 13199 X)
paperback **£5.95** (0 631 13647 9)

Basil Blackwell

AUSTRALASIAN JOURNAL OF PHILOSOPHY

Contents of Vol. 62, No. 3; September 1984

EDITOR: BRIAN ELLIS

ARTICLES

David Lewis	Putnam's Paradox
Peter J. Markie	Feinberg on Moral Rights
Stephen Hetherington	Parsons and Possible Objects
Jay E. Harker	Can There Be an Infinite Regress of Justified Beliefs
Michael H. Robins	Deviant Causal Chains and Non-Basic Action

DISCUSSION

Richard J. Hall	Pierre and the New World Makers
Albert Casullo	Conjunctive Properties Revisited
Robert Elliot and Andre Gallois	Would it Have Been Me? (Against the Necessity of Origin)
Christopher S. Hill	Watsonian Freedom and Freedom of the Will

BOOKS RECEIVED, NOTES AND NEWS, INDEX

ANNUAL SUBSCRIPTIONS: Individuals A$16 (US$18 or Stg£9); Institutions A$32, (US$36 or Stg£17) All correspondence should be addressed to the Editor or Business Manager, Department of Philosophy, La Trobe University, Bundoora, Victoria, 3083, Australia.

PUBLISHED QUARTERLY BY
THE AUSTRALASIAN ASSOCIATION OF PHILOSOPHY

MIND

A QUARTERLY REVIEW OF PHILOSOPHY

edited by D. W. Hamlyn
with the assistance of S. D. Guttenplan

Vol XCIII No 372 October 1984

Articles

The Hypothetical Imperative: R. S. DOWNIE
Are 'Scientific' Objects Coloured?: C. L. HARDIN
Does the Gettier Problem Rest on a Mistake?: R. L. KIRKHAM
Mind and Matter: A Problem that Refuses Dissolution: A. SKILLEN
The Contingent Identity of Particulars and Universals: A. CASULLO
High-minded Egoism and the Problem of Priggishness: N. M. LEMOS

Discussions

An Argument against Anti-realist Semantics: A. APPIAH
Thalberg on Inmateriality: P. GLASSEN
Frege, the Proliferation of Force, and Non-cognitivism: S. C. HURLEY
I Will, if You Will: C. RADFORD
Reply to Schueler on Akrasia: C. C. W. TAYLOR
Paul Russell on Hume's 'Reconciling Project': A. FLEW
Understanding Happiness: G. HAGBERG

Critical Notice

Wittgenstein on Rules and Private Language by S. A. Kripke: R. SCRUTON

Book Reviews **Books Received** **Notices**

Orders to: Journals Department,
Basil Blackwell,
108 Cowley Road,
Oxford OX4 1JF, England

Annual subscription:
Institutions £14.50 or $32 50 (Canada $37 75)
Individuals. £10.00 or $22.50 (Canada $26.25)
Single issues: £4.00 or $9 00 (Canada $10.50)

Published for the MIND ASSOCIATION by BASIL BLACKWELL

COPERNICUS AND THE SCIENTIFIC REVOLUTION

by Edward Rosen ORIG ED. 1984 • 224pp. • Ppb. $6 95

An ANVIL Original under the general editorship of Louis L. Snyder

Copernicus' proclamation that the earth was a natural satellite of the sun started the Scientific Revolution which has continued from his time to ours. This book presents concisely and fascinatingly the author's intensive investigations of everything that Copernicus wrote and was written about him from his own time to the present. This volume is a fresh view of Copernicus in relation to the Scientific Revolution which since his time has grown to dominate the world in which we now live.

ALSO AVAILABLE:

RATIONAL FRONTIERS OF SCIENCE by Mauricio Rocha e Silva

A PHILOSOPHY OF SCIENCE FOR PERSONALITY THEORY
by Joseph F. Rychlak

For more information, send for our descriptive catalog.

Krieger Publishing Co., Inc. P.O. Box 9542, Melbourne, FL 32902-9542

When ordering, please add $3.00 for first book ($1.00 each additional) to cover shipping.

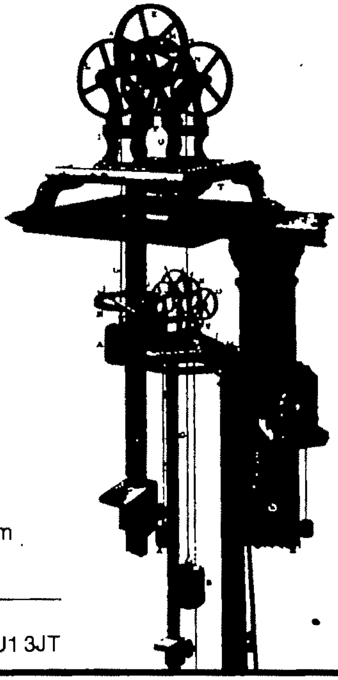
Scientific Explanation and the Causal Structure of the World

Wesley C. Salmon

Salmon describes three fundamental conceptions of scientific explanation—the epistemic, modal, and ontic, and defends a causal/mechanical theory that is a version of the ontic conception.
LPE $14.50 (U.S.) C $45.50 (U.S.)

Order from your local bookseller or from

Princeton
University Press
15A Epsom Road, Guildford, Surrey GU1 3JT



The Philosophical Quarterly

Vol.34 No.137 October 1984

Edited in the University of St. Andrews by an Editorial Board
Chairman of the Board of Editors: LESLIE STEVENSON
Executive Editor: GORDON GRAHAM *Reviews Editor:* ANTHONY ELLIS

An Unsolved Puzzle about Knowledge	*Thomas Tymoczko*
'Ought implies can' and two kinds of morality	*John Kekes*
What does the private language argument prove?	*Oswald Hanfling*
Evolutionary Epistemology and Ontological Realism	*A.J. Clark*

DISCUSSIONS

The Failure of an *A Priori* Argument for Realism	*Anthony L. Brueckner*
Ackerman on Propositional Identity	*Rod Bertolet*
Humes Dialogue IX Defended	*Donald E. Stahl*
Lucas, Godel and Astaire. a rejoinder	*J.R. Lucas*

Subscriptions for 1984

Annual Subscription	UK	Overseas	USA	Canada
Institutions	£22.50	£27.50	$50.00	$62.00
Individuals	£10.25	£12.50	$25.00	$30.00

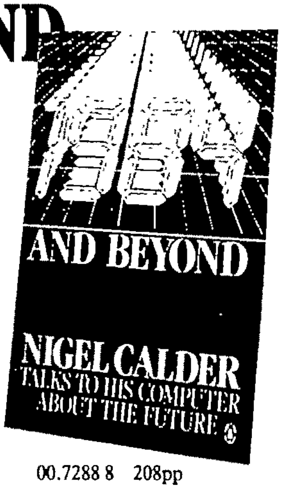
Orders, remittances and enquiries should be addressed to: Journals Department,
Basil Blackwell Publisher Ltd., 108 Cowley Road, Oxford OX4 1JF, England.

1984 AND BEYOND

Nigel Calder

- **The Computer Revolution**
- **Major exploitation of the seas**
- **The disappearance of poetry and art**

These are just some of the predictions that were made by futurologists, scientists and technologists twenty years ago for Nigel Calder's famous 1964 symposium titled "The World in 1984". Now, assisted by a supercomputer, Calder reviews the hits and misses of his project, lets the contributors comment, and begins a startling *new* cycle of predictions about 2004, from savants such as Herman Kahn, Freeman Dyson, Kurt Vonnegut, Jr., and Barbara Wootton.



00.7288 8 208pp
£2.95 December 1984

For further information please write to
The Academic Marketing
Department, Penguin Books Ltd.,
536 King's Road, London SW10 0UH

penguin'84

EDITORIAL BOARD

R. B. Braithwaite
G. Buchdahl
L. J. Cohen
Sir John Eccles
M. B. Hesse
Sir Harold Jeffreys
C. W. Kilmister
W. C. Kneale
S. Körner
D. J. O'Connor
Sir Karl Popper
H. R. Post
J. W. N. Watkins
G. J. Whitrow

EDITORIAL ADVISORY PANEL

The Editor is helped by members of the Editorial Board, members of the Committee of the British Society for the Philosophy of Science and by an Advisory Panel which includes:

Peter Clark
Jon Dorling
Arthur Fine
Allan Franklin
Ian Hacking
Colin Howson
Noretta Koertge
Hugh Mellor
David Miller
Alan Musgrave
John Worrall
Elie Zahar

THE BRITISH SOCIETY FOR THE PHILOSOPHY OF SCIENCE

Committee 1984-85

President: Dr. J. Giedymin
President-Elect: Dr. D. H. Mellor
Hon. Secretary: Dr. P. Gelpke
Hon. Treasurer: Dr. K. V. Wilkes
Members: Mr. G. M. K. Hunt, Dr. W. H. Newton-Smith, Professor H. R. Post, Mr. A. Young

The British Journal for the Philosophy of Science is the official organ of the Society, whose purpose is described by the following statement from its Constitution and Rules:

'The purpose of the Society is to study the logic, the method and the philosophy of science as well as those of the various special sciences, including the social sciences.'

The Society holds meetings, at which papers are read and discussed, once a month during the academic year. These are held, by kind permission, in University College, London. The Committee may recognise Regional Groups of Members formed to hold meetings in other parts of the country. Any Regional Group so recognised may nominate a representative Member to the Committee.

Ordinary membership is open to all interested persons on payment of the annual subscription. Membership of the Society is international and is by no means confined to people resident in the United Kingdom. The annual subscription is £4.00 for U.K. residents and £2.00 for overseas residents. Those who wish to join should write to the Hon. Secretary, Dr. P. Gelpke, Kingston College of Further Education, Kingston Hall Road, Kingston-upon-Thames, Surrey KT1 2AQ.

Brit. J. Phil. Sci. Vol. 35, No. 4, December 1984

The British Journal for the Philosophy of Science

18 JUL 1985

Contents

Articles

313–325 ROGER ARIEW The Duhem Thesis

327–343 AARON BEN-ZEEV The Passivity Assumption of the Sensation–Perception Distinction

345–358 GREGORY CURRIE Individualism and Global Supervenience

359–374 RICHARD H. SCHLAGEL A Reasonable Reply to Hume's Scepticism

Discussions

375–377 PETER GLASSEN O'Hear on an Argument of Popper's

377–380 ANTHONY O'HEAR Reply to Glassen

Review Articles

381–390 ALLAN FRANKLIN The Epistemology of Experiment (review of IAN HACKING: *Representing and Intervening*)

390–401 PETER GIBBINS Nancy Cartwright's New Philosophy of Physics (review of NANCY CARTWRIGHT: *How the Laws of Physics Lie*)

Reviews

403–415 C. W. Kilmister on K. BLACKWELL, A. BRINK, N. GRIFFIN, R. A. REMPEL and J. G. SLATER (*eds.*): *Cambridge Essays 1888–1899* by Bertrand Russell; Sarah Waterlow on EDWARD HUSSEY (*ed.* and *tr.*): *Aristotle's Physics Books III and IV*; Fred D'Agostino on DAVID LIGHTFOOT: *The Language Lottery: Toward a Biology of Grammars*, and PERE JULIA: *Explanatory Models in Linguistics: A Behavioral Perspective*; W. D. Hart on J. R. MAZE: *The Meaning of Behaviour*.

Recent Publications

Printed in Great Britain at Aberdeen University Press